

What will Eureka discover?

West European governments are warming to the French proposal for a combined venture in high technology. The motives are political, but the risks are high.

THE President of France, M. François Mitterrand, must by now be thoroughly exasperated by his European counterparts. Throughout the five years of his presidency, M. Mitterrand has been preaching the need for European collaboration in high technology. Every summit meeting of the heads of the industrialized nations has been treated to a homily on the subject, and has been met with a blend of incomprehension and scepticism. Almost as if to humour the French government, the others agreed four years ago to mount several attempts at coordination in specific fields, providing a frisson of panic for officials in advance of succeeding summit meetings, but which have not yet produced much of substance. So why should M. Mitterrand have expected more from this year's summit, at Bonn? Indeed, because the proposed collaborative programme, called Eureka, had been advertised in advance as an alternative to collaboration by West European governments in the US programme of research on ballistic missile defences, otherwise known as the Strategic Defense Initiative (SDI) or "star wars", he had every reason to expect another rebuff. That, indeed, is what happened at Bonn, but since then there has emerged a fondness for the brave concept of Eureka that is only partly diminished by the general ignorance of what Eureka is supposed to be, how it is to be organized and what it may accomplish. For the time being, the working principle seems to be that a good idea should not be crabbed simply because it is vague. More generously to all concerned, even professionally pragmatic governments such as the British now seem to be willing to consider what Eureka might have to offer (which has surprised the French).

Eureka

For fear that Eureka may get out of hand, it is important to be clear that government's motives are at best mixed. The idea of foreign collaboration in the SDI research programme seems to have cropped up first during the British prime minister's visit to Washington last December, and to have been a British suggestion whose implications were not fully appreciated at the time. Now, the US government seems to value collaboration for political reasons: governments that agree to collaborate on the research are likely to be stauncher allies in resisting the Soviet argument that even the research programme is an abomination (which it is not). Governments (Britain and West Germany, for example) at first attracted to collaboration saw two advantages: the opportunity to learn some high technology at no direct cost and the opportunity to influence the political decision that may eventually have to be made about deployment. They have since been given pause by the more careful calculation that legal difficulties would restrict the technical benefits they might hope to win and that junior partners in such a gigantic enterprise might in the end have very little influence. The French government, traditionally suspicious of US influence in Western Europe, is consistent in its unwillingness to collaborate. Politically, the interested governments should set their faces firmly against collaboration in SDI unless they can win terms that will assure their independence in the strategic arguments that must lie ahead. The more immediate question is whether Eureka is in any way the alternative it purports to be.

Eureka, essentially a French conception, is to some extent an exercise in continental (European) chauvinism. It is firmly in the

tradition of previous French technological initiatives leading to the Concorde (an uneconomic supersonic passenger aircraft), the European Airbus (a range of successful commercial aircraft), the Ariane satellite launcher (probably a success) and Superphénix (a prototype fast reactor, still unproven). Although many of these and other projects have been (or have seemed) worthwhile in themselves, French enthusiasm for them has in part derived from the wish to narrow what used to be called the technology gap between the United States and Europe (not to mention France). It is also worth remembering that some projects with this objective, such as the plan in the 1970s to beat off the US computer invasion of France, have been costly failures.

Solutions

Successful or otherwise, however, these projects have in common the circumstance that, from the point of view of Western Europe as a whole, they are second-bests. If the European Communities were not functioning as an efficient common market, and in the process generating the wealth of which Europe is potentially capable of generating, the need for the artificial definition of projects that will increase industrial company spending on research and development would have long since faded away. Instead, European companies would be spending from their own resources on research and development on a scale commensurate with that of their rivals in the United States and Japan — and the pace of innovation would presumably be enhanced commensurately. Throwing public money at joint projects must therefore be counted a stop-gap expedient for encouraging industrial development in Europe. Governments will have to worry less when there is a common market, but this will come about only when the same governments abandon their restraints on European competition, especially their failure to follow the principle that public purchases of goods and services should be open to competition. And for the time being, even (or especially) in France, national chauvinism takes precedence over European chauvinism.

Throwing money at Eureka, at least while the project is as ill-defined as at present, could be worse than a mere waste of money. As so far described, the essential characteristics of Eureka are that it should be a big project and that it should include the principal components of the SDI programme — information processing, high-powered lasers, particle beam technology, artificial intelligence and so on. For the time being, Eureka is a rag-bag of technology looking for a goal. No doubt more tangible objectives will emerge in the next few weeks. But as there is no reason to expect these to be related to market needs of any kind, the obvious risk is that Eureka will be a recipe for diverting

Biological manuscripts

Miranda Robertson, Biological Sciences Editor of *Nature*, is now based in the Washington office of *Nature*. Manuscripts offered for publication may be sent, as at present, to either the London or the Washington office; with the benefit of good communications, the speed with which manuscripts are dealt will be independent of the place at which they are received. But authors are asked please in future to send **four copies of manuscripts** intended for publication (one for each office and two for referees).



European research efforts from the more realistic goals that European companies have already set for themselves. Nobody will deny that in the course of such a programme of technological development, European technology would acquire new skills. European politicians are fond of insisting that "science for its own sake" is a waste of resources, which is debatable; what they forget is that technology, however marvellous, which is not harnessed to practical objectives has literally no purpose.

The lessons for the West European governments lately attracted by Eureka are therefore plain. Their first objective should be to sweep away the restraints on movement and competition which at present make European technology less effective than it might be. Second, they should be prepared to back with public money only those projects emerging from the discussion of Eureka that seem likely to be commercially successful as well as technically challenging. Third, they should persuade each other that technological chauvinism is not a game worth playing for its own sake, but that it may even divert scarce skill for more useful tasks. Finally, they should acknowledge among themselves (and tell their voters as much) that the technology gap, although not necessarily here to stay, cannot by its nature be quickly bridged. Impatience is a dangerous trap. □

Fraud prevention

Only colleagues can judge the value and authenticity of a person's work.

THERE is much in the belief that the pressure on people to publish research articles is part of the explanation for the cases of fraud in science that have become notorious in the past few years, but too much attention should not be paid to the suggestions from Los Angeles last week (page 447) that journals could do much to help by, for example, stamping out the tendency towards the multiple publication of the same data. Quite apart from the practical difficulties (identical manuscripts simultaneously submitted, for example), better regulation at this stage of the publication process amounts merely to the remedy of symptoms, not the underlying disease. In any case, if repeated publication helps the process of communication, by what test is that by itself a crime? Far better that reforms should aim at the abatement of the pressure to publish.

How does the pressure arise? Now that young people's promotion prospects in academic life are substantially determined by their record in research, it is natural to look for fraudsters among the rising stars. And this has partly been confirmed. But why, then, have the most conspicuous cases of fraud apparently arisen at some of the most distinguished laboratories? Perhaps work done in relative obscurity is scrutinized less carefully, while it is also possible that the pressure to seem to succeed is all the greater the more success there is about. But a further disturbing possibility suggested by last week's discussion in Los Angeles is that there may be a symbiotic relationship between those who advance their reputations by fraud and much more senior people whose reputations are apparently established. For however replete a person's resume may be with research papers, his present influence may depend to a large extent on being thought still active in research. Indeed, the success of a department or a whole school in winning research funds from outside may depend on its head's reputation as one who still beavers away in the laboratory.

In the circumstances, it is no great surprise that great men have been vulnerable to younger fraudsters. Sadly, these relationships are often indistinguishable from those between truly great men and younger talented protégés. The remedy is not easily found, for one of its components would have to be a mutual recognition by the scientific community and those among its leaders who have become inactive that it is still possible for an established scientist to run a creative laboratory even when his direct contribution to the work is small. Then, only close colleagues would be in a position to distinguish wheat from chaff — which, in any case, is the present state of affairs. □

Educational yardsticks

The British government should not attempt to measure institutional efficiency.

THE British government's empty discussion paper on the future of British higher education (see *Nature* 23 May, p.270) has predictably been condemned by academics and the organizations which represent them, but it is unusual that the chairman of the Committee of Vice-Chancellors and Principals, Dr Maurice Shock, should have so openly broken with the conventions of polite restraint. For academics and even the author of the discussion document, the Secretary of State for Education and Science, however, the most significant development in the past two weeks may be the volume of protest there has been from the House of Commons at the emptiness of the policy now made public.

A large part of the trouble of the universities in particular during the past five years has come about because they had allowed themselves to lose their political friends, once equally strong on both sides of the political spectrum. If the opposition parties now do no more than recognize that the management of higher education in Britain has become a stick with which to beat the government, higher education cannot be confident of benefiting.

But there is always a chance that the politicians will once again come to recognize the social importance of higher education, which is the only long-term guarantor of well-being.

Calculations

Meanwhile, universities and others would do well to watch out for a little-noticed appendix of the discussion document, which purports to be a first tentative attempt to calculate "performance indicators" for the objective measurement of efficiency in higher education. The idea is that there should be constructed a set of numbers measuring such characteristics of higher education as the cost of educating people in different kinds of institutions, or by means of a course of instruction, and culminating in an attempt to quantify the social value of graduates in different fields.

To be fair, the government's document emphasizes that the numbers tucked away in last month's appendix must be taken with several pinches of salt — but it then goes on to promise (or threaten) better numbers in the future. The snag, on the present showing, is that the fact that the numbers now published would not satisfy the requirements of examiners of beginning-economics courses is unlikely to diminish the influence they will have on public policy.

The most obvious defect of the calculations now attempted is that the average costs which are their chief product may be entirely misleading. It is not surprising that the average cost of educating students in British universities turns out to be higher, subject by subject, than in polytechnics, or that a medical degree is more expensive than one in engineering (which nevertheless is shown to be surprisingly higher than a science degree). But these numbers are irrelevant to the economic management of a system of higher education already in being, when the practical questions should be whether this or that change in the pattern of an institution's work would yield this or that benefit, and where the useful answers will be provided in terms of marginal, not average, costs.

All this has long since been recognized. In the present condition of British higher education, the first implication of a properly conducted study would probably be that there is too much educational plant which is under-used, or that the government could at this juncture buy a great deal of higher education cheaply, if it chose to follow that enlightened course.

Because there is every likelihood that the British government will persist with its economic analysis, and that the results will be used for a further shaping of the system as a whole, academics (and economists in particular) should do what they can to see that the job is done properly and independently. □

Nuclear testing

Size of Soviet tests disputed by geophysicists

Los Angeles

US geophysicists made angry accusations last week that the US government is "lying" over estimated yields of Soviet underground nuclear tests. Allan Lindh of the US Geological Survey said at a symposium on treaty verification organized by the American Association for the Advancement of Science (AAAS) that geophysicists agree there is no evidence that the Soviet Union has cheated on the Yield Threshold Test Ban Treaty (YTTBT), which places an upper limit of 150 kilotons on weapons tests. The Department of Defense (DoD) was accused of hiding the evidence from citizens behind a "veil of secrecy".

The treaty verification symposium was one of the more heated sessions at this year's AAAS meeting, which included discussions on several aspects of arms control. Sally Horn of DoD repeated President Reagan's charge that the Soviet Union has probably exceeded the 150-kiloton limit in the YTTBT, which went into effect in 1976. Other speakers, notably Colin Gray of the National Institute of Public Policy, argued that because the Soviets have cheated on "most if not all" arms treaties (except the Nuclear Non-proliferation Treaty), such treaties have been "harmful or irrelevant".

Although much of the information needed to judge Soviet compliance with arms treaties is subject to security classification, analyses of seismic records are published in the open literature, and have been the subject of long-running arguments. The dispute centres on the method used to estimate explosive yields from seismic data, and in particular the bias parameter incorporated to take account of the local geology of the Soviet test site at Semipalatinsk.

Yield estimates are derived from body P-waves at a frequency of about 1 Hz. The relationship between the magnitude of an event judged from P-wave amplitude and explosive yield has been thoroughly investigated at the Nevada Test Site in the southwestern United States. According to Jack Evernden of the US Geological Survey, confirmed data published more than 10 years ago (based on comparisons with other types of seismic waves) indicate that for the Semipalatinsk test site a value of 0.4 should be subtracted from magnitudes to estimate explosive yield. That judgement has recently been confirmed by, among others, the Air Force Technical Applications Center in Florida, which has a primary responsibility for yield estimates. Present US government estimates, however, use a bias value of only 0.2, resulting in yield estimates that are twice as high, as with the 0.4 value. Evernden says the decision to use a bias of 0.2 was

"obscure", and claims there is unanimity among geophysicists that the higher value is correct. When the higher value is used, yield estimates converge on the 150-kiloton legal limit.

The DoD representative at the symposium was unable to shed any light on the matter. She professed unfamiliarity with the detailed arguments but admitted there was a "debate" on the question.

Seismic verification would be a crucial

factor in any comprehensive test ban that might be negotiated in future; negotiations for such a ban were suspended in 1980. Evernden, who used to work for DoD, publicly rebutted a paper published in *Science* in January this year by Willard Hannon of Lawrence Livermore Laboratory, which questioned whether a politically plausible detection network could detect most low-yield explosions. In Evernden's view, almost all test explosions above 1 kiloton (the limit of military value) can now be distinguished from earthquakes by a simple network of non-array detection stations. Others at the conference argued, however, that the lack of progress towards a more comprehensive test ban resulted from a lack of political will rather than for technical reasons.

Tim Beardsley

US research restrictions

Pentagon prefers export control

Los Angeles

THE US Department of Defense (DoD) has told the research community that it will in future make routine use of export regulations to restrict access of foreign nationals to scientific data which, while not classified, might be of military significance. As a result, many scientific and technical conference organizers will be obliged to put sensitive material in special "export-controlled sessions" restricted to US nationals and to foreign nationals only if validated by their embassies. Participants in such sessions must undertake to keep the restricted material confidential.

The new policy was spelled out at a meeting of the American Association for the Advancement of Science by Dr Sumner Benson, substituting for Richard Perle, Assistant Secretary for Defense for international security policy. In the past, some technical societies have voluntarily restricted access to conferences for fear of violating arms traffic regulations: the new policy, by giving control to DoD, will make such voluntary censorship unnecessary.

Export controls will be used only for scientific data that are not considered "fundamental". According to a DoD policy instituted last year by Richard DeLauer, then Under-Secretary of Defense for research and engineering, "fundamental" research on university campuses is controlled only by use of security classification.

The new policy outlined by Benson would implement authority in the 1984 DoD Authorization Act allowing DoD to restrict unclassified data with military or space applications. That authority was granted because of concern that the Freedom of Information Act might allow the Soviet Union access to sensitive research results, because data released under the Freedom of Information Act cannot be controlled by export regulations.

Benson's announcement clears up much of the confusion that has been rife since DoD enacted export-controlled sessions at

an April meeting of the Society of Photo-Optical Instrumentation Engineers in Virginia. Then, it was unclear whether such sessions would become commonplace, but at a recent private meeting, organized by the Institute of Electrical and Electronic Engineers DoD apparently declared itself happy with the formula. It remains to be seen whether learned societies will be happy to accept export-controlled sessions as a matter of course.

The new policy is a further development in the administration's continuing efforts to restrict access to sensitive technology. Benson said that reliance on security classification would be expensive and cumbersome but that export-controlled sessions would be more predictable and flexible, while allowing maximum access to restricted information consistent with national security. Benson said that the new arrangements would also have the advantage of not penalizing researchers at institutions which have set their faces against classified research on-campus.

The export regulations to which DoD is appealing are themselves being revised. According to the latest draft, information resulting from research contracts which embody restraints on free publication — even non-military contracts — would require a validated export licence to be transmitted to foreign nationals.

Boyd McKelvin of General Electric Co. made a strong plea at last week's meeting for moderation in the use of export regulations. McKelvin said he was unaware of any cost/benefit analysis supporting further restriction of data flowing to friendly nations, and warned that industrial laboratory managers are concluding that they will be forced to avoid employing foreign nationals. In McKelvin's view, unilateral action by the United States to prevent access by foreign nationals to US research could backfire and lead to reprisals, causing economic damage to US industry.

Tim Beardsley

European defence programme

Favourable noises for Eureka

EUREKA, the French project for a European alternative to the American "star wars" or Strategic Defense Initiative (SDI) programme, seems to be in a much better state — politically speaking — than it was a few weeks ago. Or so French sources are making out, now that the British foreign minister, Sir Geoffrey Howe, has indicated to his French counterpart M. Roland Dumas that he favours Eureka, despite initial British reticence.

The new British position seems to be that a variable-geometry, high-technology European programme, run by what the French describe as a "light" administration that would include the European Commission but not be controlled by it, could well be effective. Sir Geoffrey, it is believed, would like Eureka to concentrate on the "coordinated exploitation" of basic research. This is a logical insistence, as it is exactly what Britain alone has for long failed to achieve.

In vitro fertilization

Travemünde

MOST participants at this year's meeting of West German physicians (*Deutscher Ärztetag*) believe that the absence of legal guidelines in modern reproductive medicine should be made good by their own initiative. Fertilizations outside the mother's womb now occur in a "legal vacuum".

The physicians resolved to adopt professional rules (*Richtlinien*) to decide between correct and incorrect behaviour. Their unanimous opinion was that such rules should not be left solely to politicians and lawyers. Although it was also clear that many physicians have no wish to play a decisive role in *in vitro* fertilization, their demand was refused that there should be a stop to the programme until the inherent complications are understood.

The final communique was described by some as a presumptuous attempt by physicians to force their morals on the rest of society, particularly the restriction of the new technology to married couples.

Supporters of the restriction, however, hold that marriage and the family are the more relevant legal and ethical values, and that exceptions should be allowed only on the recommendation of a specially founded commission. Both sides agreed that the "rent-a-mother" (*Leihmutter*) concept is unacceptable, if only because of the risks of commercialization.

The physicians have taken this initiative now because the federal government is as slow and indecisive in this context as in that of recombinant DNA technology. One aim of the meeting was to put pressure on the government to clarify the legal position so that physicians are not left to decide alone. Jürgen Neffe

However, while Paris is now happy with London's support, and while administrations in many European states are now behind Eureka, contradictory statements and confusions about the project still abound, both between nations and between senior ministers of single nations, and particularly between France and Germany. Originally, German Chancellor Helmut Kohl had embraced SDI research in collaboration with the United States while French President Francois Mitterrand had rejected it, describing it as mere "sub-contracting" and fearing a drain of technology and brains to the United States. But two weeks ago the West German foreign minister Herr Hans-Dietrich Genscher had visited the French capital and expressed doubts about star wars and enthusiasm for the Eureka project. Hopes were high for an agreement of last week's summit meeting between Kohl and Mitterrand at Lake Constance in Bavaria. In the event Kohl refused to abandon his support for SDI research, and failed to give any explicit backing to Eureka.

Afterwards, *Le Monde*, normally France's most respectable newspaper, launched an extraordinary front-page, and apparently politically inspired, attack on Kohl, describing him as provincial and vacillating, and complaining that deep issues went completely over his head. But Mitterrand on Friday lowered the temperature greatly by describing Eureka as "a Franco-German idea", by claiming that it was of "vital necessity for West Germany" and that Germany agreed with this. Thus the waters muddy, and the true Eureka is hard to find.

At least the French view is becoming sharper. For example, M. Hubert Curien, the research minister, has now clearly confined M. Mitterrand's "non" to star wars to government actions and government support, and left industry free to decide for itself whether SDI involvement would pay. In practice, then, as countries that have said "yes" or "ja" are unlikely to be enthusiastic about SDI research if it involves the spending of large amounts of government money (unless that government is American), there may be little difference between the French and other positions.

In further clarification, Curien last week made the longest public statement yet on the precise nature of Eureka. He claimed it was well received in Rome, Copenhagen and Berne as well as in London (and Bonn), and promised the first "one or two" real programmes by the end of the summer. He was at pains to emphasize the need for industrial involvement in defining programme content and objectives, and said that such consultations were well under way. As to what the programmes will be, Curien remained curiously silent.

Robert Walgate

European defence programme

Ambitions of Heriot-Watt

WHILE European governments consider how to respond to American Secretary of State Casper Weinberger's invitation to join in star wars — otherwise the Strategic Defense Initiative (or SDI) — research, the SDI organization in Washington is not letting any grass grow under its feet. Last week, an SDI representative in England visited Europe's leading laboratory in optical computing at Heriot-Watt University in Edinburgh, and all but made an offer of a \$150,000 grant, with a promise of several times that amount to come. In return Professor S. Desmond Smith and his team, who lead an eight-university European group created by the European Commission in Brussels to design the first all-optical computer, will join an American network working for SDI.

The Heriot-Watt group leads the world in this technology. In the past few months Smith and colleagues were able to demonstrate the first three-element loop of optical switches operating at visible wavelengths at room temperature. This shows that "an indefinite extension of optical logic circuits is possible" said Professor Smith on Monday. Optical computing may offer not only speed but also enormous parallel processing capability, and is of much interest to SDI proponents, as it might make possible the rapid computations needed to identify, track and destroy multiple targets.

Is this the kind of loss of European high technology to the United States that the French had in mind in putting forward the Eureka programme? Not quite. "I'm here to build Europe", says Smith, "and if this grant is not going to help them I'm not going to use it".

For Smith, the lesson of this first SDI grant in Europe, which he described as "peanuts" compared with his collaboration's existing £2 million from British and European sources over the next three years, is that the US Department of Defense seems to spend not just for defence, but for the whole of American industry. This is exactly how US industry get into new technology so fast, Smith feels. By contrast, in seeking UK support to build prototypes of a working optical parallel processing system, Smith is finding that his requirement falls neatly between the Department of Education and Science, the Department of Trade and Industry and the Ministry of Defence. But in an exchange of letters with the Prime Minister, her chief scientific adviser, Sir Robin Nicholson, and others, Smith finds there is considerable support for an improvement in interdepartmental cooperation. Meanwhile, he is keeping closely in touch with the British government about the American approaches.

Robert Walgate

Research fraud

New ways of shading truth

Los Angeles

PRESSURE to publish and a lingering "pre-med syndrome" among medical researchers contribute not only to the relatively rare cases of outright scientific fakery, but also to various much more frequent abuses of scientific publication, according to medical researchers and journal editors speaking at the annual meeting of the American Association for the Advancement of Science.

Dr Edward Huth, editor of *Annals of Internal Medicine*, said that journals and scientific societies should consider adopting a precise definition of authorship to prevent one of the most frequent abuses, "false authorship". Other common abuses that he had encountered were the practice of slicing one study into a series — "salami science" — and repetitive publication.

Huth said false authorship usually occurs when a department chairman, laboratory technician or anyone not directly responsible for the intellectual content of the paper is listed as an author. Huth told of a case in which a researcher learned only by accident that his name had been included among the authors of a paper submitted to the *Annals*. While the researcher had provided some technical advice to the main author, at another institution, he had contributed neither to the design of the experiment nor to its execution, and had not even seen the final paper or any drafts. In fact, he had specifically declined an offer to be listed as an author.

Another common deception is statistical manipulation of data. Dr John Bailar, a statistical consultant to the *New England Journal of Medicine*, said that the deliberate misuse of statistics should be part of an "expanded definition of fraud". Inappropriate statistical tests have often been used to exaggerate the significance of the results presented. In some cases, authors seem to have tried several statistical tests, reporting only that which gives the best results. Hypotheses suggested by the data are often reported with *P*-values, which have meaning only for hypotheses formulated independently of data.

Huth suggested several steps that could be taken to curtail these abuses, which, he said, because of their "frequency and ubiquity", make them "more worth our attention than fakery". Although little can be done to prevent "salami science", repetitive publication could be checked by specifically asking reviewers to bring prior publications of the same material to the attention of editors.

False authorship may be more difficult to eliminate. The *Annals* requires each author to sign a form acknowledging that he has seen the final version of the paper and agrees to its publication. The shortcomings of even this approach were revealed by the case Huth mentioned above: the

eight signatures on the forms bore a striking similarity.

Dr Robert Petersdorf, dean of medicine and vice-chancellor for health sciences at the University of California at San Diego, suggested that the problem lies deeper. Science has become "too big, too competitive, too entrepreneurial, too bent on 'winning'". Federal research grants to medical schools have grown from \$436 million in 1960 to \$2,351 million in 1981, increasing proportionately much faster than enrolment.

Petersdorf said that the "pre-med syndrome" of intense competition and resorts to cheating carries through to competition for grants and for tenure, "eroding the moral fibre" of physicians.

The deputy editor of the *New England Journal of Medicine*, Dr Marcia Angell, suggested that grant-making agencies and

tenure committees could help eliminate these abuses at source by refusing to consider more than the three articles a candidate considers to be his best in a given year.

Whether pressures to publish are responsible for the more serious cases of fraud remains unknown. Patricia Woolf, at the sociology department at Princeton University, noted that the number of papers published per scientist per year has changed little but in one group she has studied in detail (chemists), the number of "super-productive" researchers has clearly grown.

The National Institutes of Health (NIH), which investigates cases of fraud involving federal research grants, say that they receive about two reports a month of fraud or scientific misconduct, a minute fraction out of the 20,000 research projects active on any given day. Dr William Raub, deputy director for extramural research and training, said that NIH will nonetheless issue rules this summer requiring institutions that receive federal funds to report any cases of suspected fraud. **Stephen Budiansky**

Arianespace

Sky-high insurance unfair

ARIANESPACE, the largely French company that now runs Europe's Ariane space launcher on something approaching a commercial basis, is fed up with the huge insurance premiums that are being asked of satellite makers for launch on Ariane. According to M. Frédéric d'Allest, president of Arianespace, his company is being forced to pay for the losses made by insurers of the ill-fated Palapa B-2 and Westar-6 satellites, mis-launched from and then recovered the US space shuttle, and for recent failure of Syncom IV-3 and the earlier TDRS-A, also both shuttle-launched. Insurance premiums are now so high, 20 per cent of the capital cost of Arabsat, recently launched by Ariane, for example, that the growth of the satellite communications business itself is under threat.

According to Arianespace officials, before the shuttle setbacks premiums were set at around 6 per cent for a shuttle launch and 9-11 cent for Ariane. This was after Ariane had suffered two early failures, and reflected a feeling among the insurers that because the shuttle was manned, there should be less risk of uncorrectable trouble with a shuttle launch. But this did not take into account the large, multi-million dollar booster rockets needed by shuttle-launched satellites to take them from the low orbit of the shuttle up to geostationary heights, says Arianespace. This is the technology in which the shuttle has faced difficulties, while Ariane, which launches into an elliptical "geostationary transfer" orbit from which only a small "apogee motor" is needed to reach geostationary orbit, has gone from strength to strength. Ariane has now successfully launched 13 commercial satellites in its last eight flights, and it is time the insurers recognized the

fact, says d'Allest.

But the insurance companies are reckoned to have lost almost \$200 million on Palapa and Westar alone, despite their recovery, and another \$85 million for Syncom, and they are fighting shy. Premiums are now around 18 per cent, with higher figures for those the companies think can pay, such as the Arabs. The premiums are the same for the shuttle and for Ariane, claims Arianespace, and they come as a package, making no distinction between the launch to first orbit (which the shuttle has accomplished perfectly) and the boost phase. M. Deschamps, director-general of Arianespace, would like the insurers to make a distinction by charging the same to Ariane and the shuttle for the first injection, and then more to shuttle customers for the boost, thus reflecting the apparently higher risk. At present the terms are "a brake" on growth, says Deschamps.

And Arianespace is certainly looking for growth, with 25 future satellites in the queue, an order book totalling FF 6,500 million (£600 million), a second launch pad at the launch site in Kourou, French Guiana, nearing completion and a third planned, apart from French space agency work on advanced and more powerful forms of Ariane, including a mini-shuttle, Hermes. And while President Reagan is attempting to shift more of the real costs of shuttle launches onto the customers, raising launch fees, Arianespace is planning to follow suit to raise profits, but not so much as to threaten the 30 per cent of the world market for satellite launchers the company hopes to be taking from 1987 on. Here is one field in which Europe feels at last that, for the time, it is holding one or two trump cards. **Robert Walgate**

European molecular biology

Computing facilities for EMBL

Brussels

PLANS to expand databank services, increase in-house and collaborative research and improve access by researchers to software and data at the European Molecular Biology Laboratory (EMBL) in Heidelberg have been given a first airing.

The idea for a European biocomputing facility, which would give the laboratory a greater coordinating role in Europe, was proposed by its director, Lennart Philipson, and was welcomed by scientists and science policy advisers, not only from Europe but also from the United States and Japan at a meeting to discuss developments in biosequencing in Italy last month.

EMBL's biocomputing programme involves the development and application of computing methods and a large-scale public nucleotide sequence databank, set up in 1981, which exchanges information regularly with the Los Alamos databank in the United States. The preliminary proposal outlined by Philipson would mean an overall expansion of computer operations and the setting up of a software library to enable remote users from all over Europe to screen data at EMBL, and to download them, together with the software, for local analysis. A data library is also envisaged which would collaborate with other databanks to supply protein, microbiological and gene-mapping data. The aim is not,

however, to centralize all remote user computation but to act as a source of data and portable software for remote users. Computation would be mainly on a query and search basis, with EMBL providing a package of utility programmes for data screening, searching and manipulation. A software and data link-up with large national centres could also occur.

The facility would also carry out sequence analysis, developing methods and algorithms to interpret nucleic acid and protein sequences, basic research in the field of protein folding and engineering, modelling of proteins, protein-nucleic acid interactions and other more complex structures, biological graphics, image analysis and the development and maintenance of a user-oriented system and database for the analysis of two-dimensional electrophoresis. Finally, the facility could constitute a forum both for research and for the elaboration of software policy.

The expansion would require EMBL to increase staff and invest in more high-level hardware. The laboratory already has plans to install a VAX 8600 cluster core by the end of this year to which further VAX computers could easily be added. The added functions would, says Philipson, require an extension of hardware. Preliminary estimated costing of the proposal puts capital expenditure at DM 7.3 million and annual

operating expenses at DM 3.2 million.

According to Philipson, the needs could be covered by a 10 per cent increase in the current budget. At present the 12 to 14 countries participating in the laboratory (including non-EEC countries such as Sweden, Finland and Israel) contribute a total of DM 38 million to EMBL's annual budget of DM 45 million.

Extra funds for space and hardware increases could be found in two ways: by contracts involving some of the interested countries or from the European Community, says Philipson. While some countries need and welcome European cooperation in this field, others, such as the United Kingdom, have been less keen to contribute in the past, preferring to build up activities on a national level. Whether the venture will be taken up in the laboratory's second scientific programme, which will run from 1987, will depend very much on several meetings that have been planned by EMBL's Scientific Advisory Committee and, ultimately, the EMBL Council. The next meeting of the advisory committee is scheduled for three weeks' time.

Anna Lubinska

US research support

Defence money no panacea

Los Angeles

UNIVERSITIES in the United States are in for a disappointment if they expect to reap a windfall from a new Department of Defense (DoD) initiative to support academic research. The "University Research Initiative" proposed in this year's Pentagon budget is, in fact, largely a symbolic step even in the eyes of Pentagon research chief Leo Young; Young's predecessor, George Gamota of the University of Michigan, speaking at the meeting of the American Association for the Advancement of Science, noted that when DoD began a similar programme to support university instrumentation three years ago, expectations ran so high that the \$30 million a year programmes received \$600 million worth of proposals.

Young said that the new initiative would help to overcome a "perception" problem dating from the passage of the so-called Mansfield amendment in the late 1960s. This rider on the defence appropriations bill requires all DoD to research support the mission of the agency, and has often been blamed for dwindling DoD support of basic research in the 1960s and 1970s and for continuing to prevent DoD from entering into a closer partnership with univer-

sities. Gamota, however, believes that the amendment has been used as an excuse within the Pentagon for not increasing basic research funds — particularly by those who wish to see other programmes funded.

The new initiative would provide \$25 million, evenly divided between the Army, Air Force and Navy research offices and the Defense Advanced Research Projects Agency, for graduate fellowships, instrument purchase and young investigator awards. The House Armed Services Committee has recommended increasing that figure to \$200 million, but the Senate has approved the lower figure. The Pentagon is proposing a total basic research budget for next year of \$971 million. Historically, about one-half to two-thirds of that money goes to universities.

Existing programmes that support some 300 graduate fellowships, as well as the \$30 million a year instrumentation programme, would continue even without the new initiative — and despite the Mansfield amendment. Young noted that the current version of the amendment provides a wide loophole by allowing research with a "potential" for supporting the Pentagon's mission "in the opinion of the Secretary of Defense".

Stephen Budiansky

Soviet pay rises

SOVIET scientists, designers and production engineers will from next year receive higher wages and more generous bonuses. But the resolution of the Community Party Central Committee, the USSR Council of Ministers and the All-Union Trade Union Confederation, which established the new pay scales in mid-May, made it clear that the purpose of the reform is to reduce the time spent in research and development, and in the implementation of new technologies, the endemic bottle-neck of Soviet civil industry.

Under the new scheme, managers of enterprises and organizations will be empowered to pay supplementary salaries and bonuses to persons responsible for "creative contributions" to new technology. Bonuses will be more closely linked to the "economic benefit" produced by the new technologies, and several measures have been proposed to improve the system of granting "author's certificates" (the Soviet equivalent of patents) and of making the attendant payments to scientists and technologists. Proposals to establish two new honorific titles: "Meritorious Designer" and "Meritorious Production Engineer", were also approved.

Although the new pay system seems to be aimed initially at applied scientists and technologists, it will almost certainly produce some spin-off for those working in basic research. During the past decade, there has been increasing emphasis on the need to strengthen links between academic research and production, and on the coordinating role of the Academy of Sciences in ensuring that research results are quickly implemented.

Vera Rich

University of California

Oil disquiet at Santa Barbara

Los Angeles

THE University of California at Santa Barbara (UCSB) is about to build 65 new faculty houses with spectacular views of the Pacific Ocean, nearby beachfront and modern conveniences, all within walking distance of campus. But university officials are worried that professors attracted to their seaside campus might be disappointed. The reason is that UCSB has rich oil deposits in its front yard. An oil company plans to erect two giant drilling platforms in shallow water only a mile and a half away from campus and from the new faculty housing. And it is not just the view that is being threatened: the university's marine scientists fear that a marine study-preserve in the waters off campus could be harmed by the mud produced in drilling.

UCSB chancellor Robert A. Huttenback seems reconciled to development of some kind, but says that the way it is done may be crucial. "If the air gets mucked up, the research environment could be ruined. The university could lose its appeal to students

and faculty."

Everyone agrees the project will go ahead. The state of California is eager to lease the offshore tract to Atlantic Richfield (Arco), a Los Angeles-based oil company, because the state expects to earn between \$600,000 and \$1 million a day when the facility reaches peak production in 1992.

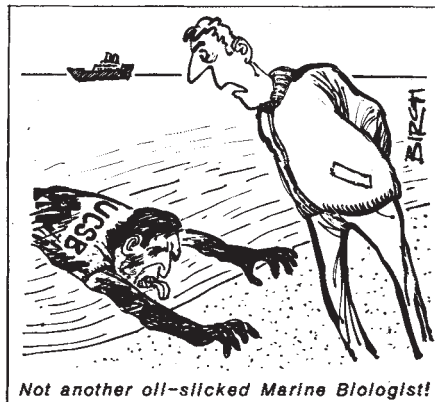
Arco has held its lease on the tract since 1945; but has only recently discovered the major new oil deposits. The Santa Barbara Channel and nearby Santa Maria Basin, most of which are under federal control, are expected to become the second largest offshore oil development in the Western hemisphere. Six other oil companies hold leases off Santa Barbara's coast but only Arco's is in the university's front yard.

The project calls for two double platforms to rise about 14 stories above the ocean floor and will produce from 80,000 to 500,000 barrels of oil a day. Plans for treating and transporting the crude oil and natural gas are under discussion.

On campus, likely winners include the

structural and ocean engineers who will participate in designing new offshore platforms. Most worried are the marine scientists. Those affected by the project bring in over \$5 million a year to the university in outside grants.

Aquatic biology has 1,450 undergraduate majors and 148 graduate students, making it the nation's largest marine sciences



programme for undergraduates. Pioneering work has been done on abalone culture, bioluminescence, kelp reproduction and sediment transport. A specialized seawater system pumps water into laboratories and classrooms.

Because of their stake in the waters off campus, UCSB scientists are trying to convince state authorities that the usual legal requirements for protecting the environment are not enough to safeguard the area for research purposes.

Will toxic drilling muds damage the marine habitat? A preliminary study, paid for by the chancellor's office, indicates that such muds are detrimental to juvenile but not adult organisms. There is worry the muds might kill young abalone and other species and also foul the seawater intake system. To quell such fears, Arco has said it might barge drilling wastes ashore to avoid contamination.

The university also fears there will be a housing shortage caused by the influx of workers. Rents are already high and the school will have to increase financial aid so that students can afford to live nearby.

But aesthetic concerns top the list. The campus is only two miles from the Santa Barbara airport and some professors are already bothered by helicopters taking off every 15 minutes or so as part of the overall oil development taking place in the region. Oil platforms, no matter how well built, strike many people as ugly. And their operations do sometimes emit foul-smelling odours that even now (from other oil rigs seven or more miles offshore) waft across campus a couple of times a month.

The views of the university are important to Arco, according to Robert Flaherty, a company vice-president. Nobody in the industry would produce oil and gas at the price of damaging UCSB's educational and research activities, he says. "We can't see the remotest possibility of these terrible effects."

Sandra Blakeslee

Franco-Japanese voyage

Hitch mars start for *Nadir*

Tokyo

NO sooner had the French research vessel *Nadir*, with the deep-sea submersible *Nautilie* riding piggy-back, set sail on 1 June on the second phase of the Franco-Japanese deep-sea project Kaiko than mis-hap struck with an accident that injured two French crewmen, one of them seriously. They were struck by flying metal from a ring that broke when testing the crane used to lower *Nautilie* from the mother ship. Two relief crew are being flown out to Japan. Once they are on board, the *Nadir* will immediately set sail again on its three-leg cruise to explore the plate boundaries running along the deep-sea troughs and trenches off the coast of Japan.

The accident aboard the *Nadir* must add to the worries of Kazuo Kobayashi, Japanese co-director of the project, whose worst nightmare is that the *Nautilie* will collide with a ship when she surfaces. The seas off the Pacific coast of Japan are criss-crossed by some of the busiest shipping lanes in the world and, near the junction of the Japan and Kuril trenches, where the cold southward-flowing Oyashio current meets warm air masses from the south, fog is frequent. As the *Nautilie* has a cruising speed of only one knot, extensive surveys of currents in the study areas have been carried out before the survey to avoid unnecessary "drift" of the submersible during its planned two-hour descent to 6,000 m. Transponders placed on the bottom will allow navigation with pinpoint accuracy and the submersible can be directed by telephone to spots of interest by the mother

ship. Time on the bottom will be about five hours at 6,000 m during which the submersible can traverse about 9 km of sea floor. The three-man submersible will carry only one scientist at a time, so French and Japanese researchers will take turns to descend.

The temporarily delayed voyage is due to finish on 19 June after exploring the Tenryu Canyon and Zenisu Ridge which flank the Nankai Trough — a sediment-filled basin that straddles the boundary between the Eurasian plate on which Japan rides and the subducting Philippine plate which plunges beneath Japan.

The second leg, from 22 June to 10 July, will traverse the Suruga and Sagami troughs that cut into the Japanese mainland at the leading edge of the subducting Philippine plate. *Nautilie* will then investigate Dai-ichi Kashima Kaizan, a Fuji-sized seamount that is being sliced and swallowed up by the Japan Trench (*Nature* 313, 433; 1985). On the final leg, from 13 July to 1 August, a seamount in a similar predicament will be investigated at the junction of the Japan and Kuril trenches, and it is hoped to make the deepest dive to nearly 6,000 m.

Nautilie may also find subduction-zone analogues of the "vent communities" of giant tube worms and clams found at spreading centres, but here associated with cold seeps of methane rather than hot sulphur-rich waters (see *Nature* 313, 339; 1985). No biologists, however, will participate in the dives, which will be manned by geophysicists, geologists and geochemists.

David Swinbanks

Polish education

Threat of repressive new law

THE current clampdown on the academic community in Poland could seriously imperil Polish hopes of regaining Most Favoured Nation status (MFN) with respect to the United States. Loss of MFN, as part of the package of sanctions imposed by the United States when General Jaruzelski imposed Martial Law in December 1981, has cost an estimated \$50 million in Poland's foreign trade earnings. Alleged government plans for the deployment of graduates, recently described by the underground Solidarity journal *Tygodnik Mazowsze* could well render this loss permanent.

According to *Tygodnik Mazowsze*, new legislation is being drafted jointly by the Ministry of Science and Higher Education and the Ministry of Education to ensure that all "graduates" work for ten years in the nationalized sector of the economy. "Graduates", the paper explains, apparently quoting from a leaked draft of the proposed legislation, includes all "Polish citizens who have completed basic vocational schooling, secondary vocational schooling of all types, higher education or who have taken PhDs abroad". If such persons fail to work in the nationalized sector for at least ten years after graduation, they may be asked to return to the treasury the cost of their education. As a corollary, "graduates" who have not completed this ten years' service, if going abroad, will be required to deposit the cost of their education as a guarantee of their return.

The last requirement appears to contravene the terms laid down in the US Senate by the Jackson-Vanik amendment (law 504A) of 1974, under which the US government is forbidden to grant MFN to any other government which requires intending emigrants to repay the cost of their education. According to the extracts of the draft

quoted in *Tygodnik Mazowsze*, implementation of the proposed law is expected to bring into the treasury several hundred million zloty a year from the contributions of graduates to the nationalized sector, as well as forfeited "guarantees".

In addition to the MFN problem, the implementation of the law could seriously damage Poland's international image. The new law would apply for 15 years after the end of their studies. This would seriously hamper the prospects of younger Polish scholars wishing to travel abroad for, even if they intended to return, they might not be able to pay the deposit. Linkage of higher education with employment contravenes the international human rights law.

The extension of the term "graduate" to cover the trade-school sector could hit not only small craftsmen, but also the run-down private agricultural sector. Since 1980, the Polish government has given various grudging guarantees to the private farmers (who account for some 75 per cent of all Polish agriculture) that no attempts will be made to force them into the system of state and collective farms. The government has reluctantly agreed the principle of a scheme, proposed by the Catholic Church, to raise agricultural capital abroad and, according to Fr Alojzy Orszulik, the press spokesman of the Polish Catholic hierarchy, the government negotiators are about to "finish their work" on the plan. The rumoured new law would, however, mean that graduates of agricultural academies and agricultural secondary schools could return to their family's farms only in middle life. The foreign patrons of the Church scheme (who include the governments of the United States and the European Communities) may well be given pause.

Vera Rich

Failure for West German Greens

LAST month's election in Nordrhein-Westfalen (NRW), the most populous state in the Federal Republic, was a great success for the Social Democratic Party (SPD). It comes exactly halfway through the legislative period of the Christian Democratic (CDU) Liberal (FDP) coalition government in Bonn and is widely regarded as a vote against the coalition policies. Equally important was the failure of the Green Party to obtain the 5 per cent vote necessary to win seats in the NRW parliament.

The Green Party fared less well than it had hoped in the election in Saarland three months earlier, where it obtained only 2.5 per cent of the votes. The 4.6 per cent in NRW was not a significant improvement on the result five years ago, when the Greens were a completely unknown factor.

Coalition with the SPD was made impossible by the Greens' determination that

NRW should withdraw completely from the nuclear industry. But, like the CDU/FDP in Bonn, the SPD in NRW has a strong interest in maintaining, among other nuclear power stations, the nuclear research centre at Jülich and the fast breeder reactor at Kalkar. It seems the SPD has now adopted enough of the political standpoint of the Green Party to weaken its appeal to ecologically minded voters.

Immediately after the election, an old quarrel broke out within the Green Party between the fundamentalists and realists. The former oppose all political compromises, while the latter strive to exert direct political influence. The difference between the two factions is so basic that it may splinter the Green Party. The three established parties may henceforth need take less notice of new radical ideas.

Jürgen Neffe

Indian cooking

Chulhas win fair wind

New Delhi

CHULHAS (wood stoves) that do not emit smoke and use less wood are fast replacing traditional cooking stoves in rural households, ushering in a "smokeless revolution" in India's countryside.

Some 800,000 smokeless chulhas have so far been installed and 5,000 villages have become smokeless in 15 months since the project was started by the Department of Non-conventional Energy Sources (DNES). The project has exceeded expectations, says DNES secretary Maheshwar Dayal, whose department has fixed a target of five million improved cooking stoves by 1990.

About 40 per cent of all the energy in India is used for cooking food. The traditional woodstoves made of mud and brick in India's 112 million homes annually burn 133 million tonnes of firewood and over 200 million tonnes of cow-dung cake. But the stoves are so inefficient that rural women spend long hours collecting firewood. The improved cooking stoves, with a combustion efficiency of more than 15 per cent, have drastically cut down consumption of firewood. And because these stoves are designed to burn the fuel completely, the kitchens are smoke-free. Dayal says that smokeless chulhas introduced have saved 600,000 tonnes of firewood. By 1990, the saving would be 4.2 million tonnes, valued at Rs 2,940 million (£210 million).

Some 29 models of improved cooking stoves have been developed and introduced by social organizations and research institutions such as the Indian Institute of Technology (IIT) in Delhi, the Central Power Research Institute (CPRI) and the Indian Institute of Science (IIS) in Bangalore. The cooking stoves vary according to the region, type of food cooked and fuel used. But all new designs are tested and certified by testing and training centres at IIT, CPRI and the Punjab University of Agriculture in Chandigarh.

While fixed-type chulhas are installed by government agencies and voluntary organizations, portable models are marketed at prices ranging from Rs 38 to Rs 63 (£3 to £5) by 12 private companies that have sprung up to cash in on the smokeless revolution. As an incentive to buyers, the government has abolished excise duty on fuel-efficient wood-burning stoves.

DNES has also developed community chulhas capable of cooking for up to 200 persons, with fuel efficiency as high as 35 per cent, recommended for hostels, hotels and canteens. The Tamil Nadu state government is planning to install the improved community chulhas in the 33,000 schools whose midday meal project is being threatened by firewood shortage.

K.S. Jayaraman

Oceanography

Sino-Japanese collaboration

Tokyo

IN the wake of successful cooperation between the United States and Japan in oceanography, Japan's Science and Technology Agency has recently revealed plans for joint research with China in a multi-million dollar long-term study of the East China Sea and the Kuroshio, a huge warm ocean current in the Western Pacific that affects fishing, weather and shipping.

Although details are still under negotiation, the project will start in March 1986 and run for seven years. The chief participants will be Japan's fishery, meteorological and maritime safety agencies and China's National Bureau of Oceanography.

The project, which was proposed by China in "vague" terms two years ago and put into more concrete form by Japan, will follow on from a 10-year study of the Kuroshio sponsored by the Science and Technology Agency. That survey, the Kuroshio Exploitation and Utilization Research (KER) project, is due to end next year but has only covered a small area of the East China Sea, a politically sensitive region for oceanographic research because of its potential oil reserves.

The Kuroshio (black stream, black because of its low reflectivity) is a 100-km wide "river" of clear warm water originating east of the Philippines that flows north past the eastern coast of Taiwan and then winds its way along Japan's Pacific coast before turning abruptly out into the Pacific on meeting the southward-flowing cold Oyashio current. At this bend in the Kuroshio, the Kuroshio Extension, huge swirling eddies of warm water extending to depths of hundreds of metres break off from the main stream and head slowly north, often persisting for periods of more than a year before finally being absorbed into the Oyashio or re-attached to the Kuroshio.

With a speed of 3-4 knots and a volume transport of 30-60 tons per second the Kuroshio represents a tremendous potential source of energy, a point not missed by energy-hungry Japan; one purpose of the KER project is to assess the viability of harnessing the Kuroshio's power. But the real power of the Kuroshio lies in its influence over Japan's fisheries.

In the East China Sea, the current creates one of the world's largest fishing grounds and is an important spawning area for several commercial fish such as sardines. According to a KER report published in March, vast numbers of fish eggs are spawned next to the temperature/salinity front of the Kuroshio and eggs and larvae are picked up by the current and transported hundreds of kilometres downstream to nursery grounds off the Pacific coast of Japan. In addition, fish such as tuna and bonito feed along the length of the Kuroshio and the warm eddies north of the

Kuroshio Extension can cause switches in the migration routes of several commercial fish such as Pacific saury. Japan's vast fishing industry thus has a tremendous vested interest in understanding the vicissitudes of the Kuroshio. In fact some large tuna fishing vessels are already equipped with automatic picture transmission equipment for receiving and processing real-time infrared satellite images of the Kuroshio. As such equipment is beyond the resources of all but the largest fishing vessels (a terminal costs about \$65,000) the Japan Weather Association (JWA) introduced a satellite data service in October last year which provides HRPT (high resolution picture transmission) infrared imagery data in photographic or floppy disk form for as little as \$4.60.

Satellite-monitoring both of drifting buoys and sea-surface temperature will also

form an important part of the joint project with China, according to Dr Hideo Nishida of the Hydrographic Department of the Maritime Safety Agency. The annual budget for Japan's contribution to the project will be similar to that of KER, he said, more than ¥100 million (\$0.4 million).

At the same time as the Science and Technology Agency's project gets under way in the East China Sea, the Ministry of Education-affiliated Ocean Research Institute (ORI) of the University of Tokyo intends to tackle the other end of the Kuroshio, focusing attention on the warm core rings north of the Kuroshio Extension. Professor Takahisa Nemoto, head of the Plankton Division at ORI, announced at an international seminar here in Tokyo this week that a cruise of the institute's RV *Hakuho Maru* will study Kuroshio warm core rings in September 1987, while the institute's smaller research vessel, the *Tansei Maru*, will carry out preliminary studies in the same area in 1986.

David Swinbanks

France

Future of encyclopaedia in doubt

THE creation of a great new French encyclopaedia, mostly of the sciences, was announced by the present French government in 1982. It was to use all available new technologies (databanks, electronic access, video) as well as the more traditional ones (print) and it was to be the 20th century's answer to the *Encyclopédie Diderot*, the great enterprise of the Enlightenment to encompass all knowledge in a few fat volumes. Where it is now? Appealing, it seems, to the French public for a few sous. A new advertisement has appeared in Paris asking for "patrons" to invest in the new encyclopaedia, which, until recently at least, has been unable to attract a French publisher willing to carry any financial risk.

The publishers North-Holland, it is hoped in Paris, may produce an English version, but the *Commission Diderot*, which is in charge of the 20th century masterwork, needs to raise another FF 10 million (around £1 million) to bring this "encyclopaedia of living questions" to the French. It will be interesting to see if the French will be willing to pay. So far the commission has a grant of some FF 750,000 from the ministries of education, research and culture, and has been promised a further FF 4 million from industry. That not being enough, the appeal appears to have been the only way forward.

So will the encyclopaedia, the brain-child of present education minister Jean-Pierre Chevènement (who was research minister at its launch), be worth a minimum of FF 100 investment? The French market is not large and the publishers' judgement should be respected. But FF 100 to be associated with a great project?

The encyclopaedia is certainly great: 200 volumes are expected by the end of 1989,

and the first 20, involving 200 authors, are promised by April 1986. But its structure is novel and even contradictory. The principle is that "knowledge" is not fixed but changing; and moreover that at any time there can be many opinions on a subject. While this may be true of many ideas, there are also many of which it is not true. And it is difficult to see how a flux is to be fixed in encyclopaedic form.

In practice, the *Commission Diderot's* strategy was to send out a nation-wide "appeal for questions (this came before the appeal for funds), an appeal which led to a list of 350 (such as "will carbon-dioxide emissions alter the climate and by how much?") out of which the commission has selected some 50 for the first volumes. Some 80 per cent of the questions received were in the physical or biological sciences or in mathematics, and only 20 per cent in the social sciences (which surprised the commission secretariat). Answers to these questions by the 200 authors will form a resource, it is said in Paris, from which many kinds of publication may be drawn. A great deal depends, therefore, on these authors, who must produce texts which are "intelligible to all" and address all the fears and hopes raised by the subjects they cover.

According to Dominique Lecourt, shortly after he was appointed director of the project, the encyclopaedia is to be "a powerful instrument of liberation". It will also "reject any attempt to assemble the sum of all knowledge or to present it according to an apparently rational order". It is a mountainous and confusing task, and it must be asked if even France is large enough to face it — even with a few more sous.

Robert Walgate

Lobbying of NSF denied

SIR — In "US engineering research: Six special centres founded", (*Nature* 11 April, p.488; 1985), Tim Beardsley reports on several grants made by the National Science Foundation (NSF) to establish major engineering research centres. At the end of the article, Beardsley implies that political influence was involved in the selection of one of the awardee institutions. This letter's erroneous implication could cause great potential harm.

Beardsley states: "At least one US senator had lobbied NSF to locate an engineering research centre in his state, apparently successfully. But Bloch [Erich Bloch, director of NSF] insists that siting decisions had been made exclusively on the basis of peer review of the 142 proposals from 106 institutions."

The only communication about this matter that NSF received from the senator concerned was a form letter asking for information about the proposal on behalf of a constituent. That would hardly seem to constitute "lobbying" even by the most rigid of definitions.

Beardsley apparently reached his conclusion after reading a press release issued by the senator's office that seemed to imply a greater involvement. Other writers attending the press conference about the awards correctly identified this as a case of a political figure being overzealously associated with an event beneficial to his constituency. Beardsley's statement beginning, "But Bloch insists..." does not help matters greatly.

NSF goes to great lengths to ensure that decisions on all proposals are made on the basis of scientific merit and that the scientific and engineering communities are aware of and confident in that fact. In the case of the engineering research centre proposals, the review involved a four-stage process in which a number of competent persons participated, including highly respected engineers from academia and industry. Let me state clearly that politics was not and is not part of the process. The implication in the Beardsley article to the contrary is most unfortunate.

RAYMOND E. BYE JR

Office of Legislative
and Public Affairs,
National Science Foundation,
Washington, DC 20550, USA

Tim Beardsley writes: While there is no evidence that NSF's reviewing was influenced by political pressure, the form letter referred to, from Senator Alfonse D'Amato of New York, was sent to NSF in response to a request for support from Mr Gregory Fusco, vice-president of government relations at Columbia University. The senator's involvement was readily confirmed by his office at the time the story was written. Senator D'Amato asked NSF to keep him informed of the

progress of Columbia University's application and asked that Mr Fusco's request "be given every consideration". Some might disagree with Mr Bye's assertion that this does not constitute lobbying (*Oxford English Dictionary*: "to seek to influence"). Only the over-sensitive, however, would object to the word "insists" in connection with Bloch's denial that political factors influenced the selection process; insisting that one is innocent carries no implication that the opposite is true.

Embryo research

SIR — In their attempt to establish a status for the *in vitro* fertilization embryo, which denies the scientific understanding of the human embryo and the medical code of ethics in regard to it, Drs Evans and McLaren make some remarkable points¹.

They castigate Professor Lejeune because he "... omits to say that research on the detection and prevention of these genetic defects... in some instances can only be done at this early (embryonic) stage" while omitting to point out themselves that most developmental genetic research work can be and is being done on ethically suitable animal and insect models. Indeed, *Nature* is organizing an international conference ("Genes and Systems in Development", San Francisco 7-9 October 1985) "to address what is perhaps the central unanswered question of biology today: How is the process of development directed by genes?" Most of the papers are based on animal models.

When the normal developmental processes of pattern formation are understood, then the abnormal processes can be solved — for how can one hope to understand the pathological without a proper understanding of the normal? In regard to the use of insect embryo models for understanding genetic development, David Ish Horowicz has suggested a relatively short time span: "It is conceivable that within 5 years it will all be over. We will have solved the universal biological problem of pattern formation"². Allowing a further five years to solve the question of abnormal pattern formation (genetic defects), the answers should be available before the turn of the century. In the meantime, we can console ourselves in the knowledge that infants with genetic defects comprise less than 1 per cent of live births and 993 out of every 1,000 embryos, which implant with genetic defects, will not result in a live birth⁴.

PATRICK W. GILL

89 Davies Road,
Claremont, Perth,
Western Australia 6010

1. Evans, H.J. & McLaren, A. *Nature* 314, 127-128 (1985).
 2. Moore K.L. *The Developing Human*, 8 (Saunders, Philadelphia, 1973).
 3. Vines G. *New Sci.* No. 1483, 30 (1985).
 4. Schlesselman J.J. *Am. J. Obstet. Gynecol.* 135, 135 (1979).
- ©1985 Nature Publishing Group

Women students

SIR — In his article, "Women in science and engineering", (*Nature* 2 May, p.84), Richard Pearson suggests that the cause of the surprisingly low representation in the "less traditionally heavy engineering" field of electrical engineering and electronics may be that "the competition for places is much harder in these subjects".

Surely this implies that women engineers are simply not as able as men in the field of electrical engineering and electronics, and therefore cannot compete for these jobs?

I find this an extraordinary suggestion from a member of the Institute of Manpower Studies, but comparable to the engineering careers consultant at my old university who advised all female engineers to become accountants.

ELIZABETH FISHER

42 Salisbury Avenue,
Cheam, Surrey SM1 2DJ, UK

SIR — In Richard Pearson's article on sexual discrimination (*Nature* 2 May, p.84) there is an important difference between what is purported to be the percentage of medical students in the United Kingdom who are women, and what the government believes the figure to be. Pearson's Table 1 gives 49 per cent as the percentage of first-year women medical students in 1982 (a figure he attributes to the Universities Central Council on Admissions). The Department of Health and Social Security publication *On the state of the public health for the year 1983* (HMSO) gives the percentage number of women entrants in UK medical schools in 1982-83 as 44 per cent, and for the previous year 43 per cent. Whom do we believe? Medicine being what it is, I imagine the DHSS figure is nearer the mark.

JOE COLLIER

St George's Hospital Medical School,
Department of Pharmacology,
Jenner Wing, Cranmer Terrace,
London SW17 0RE, UK

Liberation all round

SIR — We were appalled to read the report from R.A. Tatouille (*Nature* 28 March, p.323) on the formation of the Vegetable Liberation Front. This is elementism of the worst kind. We have formed the Association for the Liberation of Long Oppressed Ingots to counter this oppression, and to lobby support for non-carbon based objects.

Remember the rallying cry of ALLOI "Metals of the World unite — you have nothing to lose but your purity".

GUY A. BROWN
W.H. SWALLOW

DSIR Ilam Research Centre,
27 Creyke Road,
PO Box 29-181,
Christchurch,
New Zealand

Avoiding recurrent catastrophes

The death of perhaps 15,000 people in Bangladesh last month was caused by cyclonic and tidal forces in concert. Prevention may be impossible, but better warning is essential.

THE tragic loss of life in the Ganges delta of Bangladesh two weeks ago will never be accurately quantified. Even in normal circumstances, the numbers of people eking out a living in the delta can only be guessed at; the numbers of those swept away by the tidal wave that filled the Bay of Bengal on 25 and 26 May will be counted only imperfectly. But tragedies like this recur at frequent intervals along this low-lying coast. What can be done to prevent them, or at least to mitigate their effects? Understanding the phenomenon must be a first step, which is why there is interest in a calculation of the interaction between wind-driven surges and the natural tide in the Bay of Bengal, published in the *Philosophical Transactions of the Royal Society* (A313, 507-535; 1985) by B. Johns of the University of Reading and associates from the Indian Institute of Technology at New Delhi: A.D. Rao, S.K. Dube and P.C. Sinha.

In the long run, accurate prediction of these tidal waves must be a prerequisite for avoidance action. Two weeks ago, the meteorological services did warn of impending trouble along the low-lying coast, although the argument continues whether the warnings might have been fuller. Too fatalistically, people also ask what value broadcast warnings may be to people too poor to own radio receivers.

The essence of the phenomenon that killed about 15,000 people in the Bay of Bengal is that by which a cyclonic pressure may create a surge of water elevation as it moves through shallow water onto the shore of an enclosed tidal basin. The first expectation is that the increase of water level and thus the risk of flooding will be greater when the arrival of a cyclone coincides with high tide. But by now it is accepted that the interaction between the tides and the wind-driven surge is not linear. The whole may be greater (and more damaging) than the sum of the parts.

Johns *et al.* have provided a valuable snapshot of where the modelling of tide/surge interactions now stands, and also of what remains to be done. The paper sets out to model retrospectively the tidal wave that hit the low-lying coastal strip of the Indian state of Orissa, on the west of the Bay of Bengal on 3 and 4 June 1982. When more is known of the track of the cyclone that drove last month's tidal wave onto the coast of Bangladesh, the same mathematical apparatus may be used to reconstruct that sequence of events as well.

The technique is in principle straight-

forward. First, the bottom topography and the coastline of the bay are represented by smoothed curves. This is the idealized but still realistic basin in which the water is driven by the forcing of tidal oscillations in the Indian Ocean, and by the surface winds of cyclones. But because of the large volume of water in the several mouths of the Ganges, one elaboration of the model puts a series of 200-km stretches of water at the head of the bay.

The hydrodynamic equations for the motion of the water in the bay are simplified to the case of a flat surface, although the Coriolis forces associated with rotation are allowed for. The coupling between the Indian Ocean and the bay is supposed to be determined primarily by the astronomical tides, whence arises the most immediate difficulty of the model: by what conditions should the interaction be represented on the boundary, taken as an imaginary line running east from the Sri Lankan coast? In principle, if there were enough tidal stations around the bay, it would be possible to determine this external forcing function accurately, simply by running the model and comparing its predictions of the tides with observations of them in normal circumstances. In the event, the data are so sparse that Johns *et al.* can do no more than represent the tidal effect of the Indian Ocean on the bay by means of a regular sinusoidal variation of height with constant phase along the whole length of the arbitrary boundary. If catastrophic surges in the bay are to be predicted accurately, this needless simplification must be avoided, from which it follows that much better tidal data must be collected from many more stations in the bay.

Incorporating cyclones in the model is also an empirical process. The chief physical effect is that of cyclonic winds on the sea surface. Johns *et al.* represent the Orissa cyclone by a wind-field modelled to allow for decreasing velocity with increasing distance from the centre, but fitted with the crude meteorological data available for the cyclone. By their account, the predicted height of an induced surge in the Bay of Bengal will depend sensitively on the wind-field pattern out to 100 km or so from the centre of the storm, which points to another obvious difficulty: because there is at present no way of inferring precise wind-speeds from synoptic meteorological data, real-time predictions are, for the time being, uncertain.

In the circumstances, it is remarkable

that the model functions as well as it does. By forcing the movement of the mass in the bay by the crude representation of the effects of the external tide, as well as by the Orissa cyclone following its measured track, Johns *et al.* conclude that the average height of the tidal wave in June 1982 may indeed have reached 8 m at some parts of the coastline. But given the unavoidable arbitrariness with which the constants specifying the wind-field must be chosen, the real proof of the pudding is that the model seems to have reconstructed the sequence of events correctly.

The Orissa inundations of 1982 turn out to have been interesting in their own right (but perhaps they all are). According to the model calculations, not much seems to have happened to the level of water off the Orissa coast (except for a gradual rise totalling 2 m in a day) until the centre of the cyclone made its landfall on the afternoon of 3 June. Thereafter, however, the computer calculations bear out such anecdotal evidence as there is that the largest elevation of sea-level occurred some hours later, further to the north-east. During this interval, the model shows, it is almost as if a tidal wave was moving steadily along the coast from the point of landfall, at first increasing with height. The tidal and cyclonic effects are not additive, which is what would be expected of solutions to a set of equations which are inherently nonlinear, but the result is that some odd things can be expected. The times of maximum inundation may differ by several hours from those of high tide. Parts of the coast far from the point of landfall may be inundated twice.

Directly, none of this has much to do with last month's tragedy. Indeed, the physical scale of these phenomena is so great that outright avoidance is clearly beyond reach. The only realistic way of mitigating recurrent damage on last month's scale is that there should be more accurate warnings, and that better use should be made of them. Much might be done by running the model of Johns *et al.* not merely on those cyclones which turn out to be damaging, but on the much larger number of meteorological depressions in the Bay of Bengal which turn out to be relatively benign, for then it might be possible to build up a kind of taxonomy of dangerous cyclones. But for such an exercise to be worthwhile, better tidal observations are essential.

John Maddox

Neurosciences

Modulation of some deadly sins

from Charles F. Stevens

SEVERAL recent articles, three of them elsewhere in this issue, shed light on neuromodulation, the shadowy side of brain function. We understand quite well in terms of cellular mechanisms about half of what the brain does — the integration and transmission of information. But the other half — the neuromodulatory systems that seem to mediate, for example, gluttony, sloth, lust and anger, if not more of the seven deadly sins — is only now beginning to be elucidated.

The well understood part of brain function has been revealed through a line of research initiated by Hodgkin, Huxley, Katz, Kuffler and Eccles. Integral membrane proteins, known as channels, control the permeability of the neuronal membrane to ions, and thus regulate the brain's electrical activity. The channels change their conformation in response to synaptically released neurotransmitters or variations in membrane voltage. These conformational changes cause the channels' transmembrane pores to open and permit ionic current to flow across the neuronal membrane. The neurotransmitter, the first messenger, thus triggers the current flow that acts as a second messenger by producing changes in membrane voltage in remote parts of the neurone, where the behaviour of other channels is in turn altered. The number of distinct kinds of channels is quite large: neurones have at least a dozen species of channels that sense membrane voltage, and more than a dozen types that open their pores after binding specific neurotransmitter molecules such as acetylcholine, dopamine or serotonin.

Neurotransmitters are released at the brain's 10^{13} synapses, and part of their function is to cause channels to open. But some neurotransmitters also have a quite different function that is not yet fully understood. They bind to receptors on the neuronal surface and employ chemical second messenger systems, rather than current flow, to produce myriad effects. A typical second messenger is cyclic AMP, produced by neurotransmitter-activated adenylate cyclase. Second messengers in the brain have a wide variety of effects, ranging from the ability to alter the rate of transcription of specific genes (see, for example, Barinaga, M. *et al.* *Nature* **314**, 279; 1985) to the modification of channel properties, as described below. The hallmark of these modulatory effects is that they involve covalent modification of proteins rather than simple conformational changes, and thus occur on a time scale of seconds to minutes or longer, whereas the typical channel opening and closing that is responsible for the nerve impulse, for example, occurs in the order of milliseconds.

The three papers on neuromodulation in this issue involve three different types of channels that respond to membrane voltage by opening pores selective for potassium ions; more than half a dozen distinct types of potassium channels are known. On page 501, Aghajanian reports that the neurotransmitter noradrenaline causes the suppression of one species of potassium channel ('A' channels) in neurones of the dorsal raphe nucleus. Because the normal function of this channel is to slow the production of nerve impulses, the suppression results in an increase in the rate of firing. Stanfield, Nakajima and Yamaguchi report on page 498 that the neuropeptide substance P suppresses a second type of potassium channel (the inward rectifier), so increasing the excitability of neurones cultured from the globus pallidus. Finally, on page 503, Ewald, Williams and Levitan demonstrate the mechanism by which cAMP increases the activity of a third type of potassium channel (the calcium-activated potassium channel). Using patch clamping of isolated membrane patches from nerve cells of the snail *Helix* and channel reconstitution in artificial lipid bilayers, they find that the action of the catalytic subunit of cAMP-dependent protein kinase increases the probability that the channel will open. Their observations demonstrate that the substrate of the kinase is either the channel itself or a closely associated entity. Similar experiments by

Shuster *et al.* (*Nature* **313**, 392; 1985) have shown that the cAMP-dependent protein kinase can act to decrease the activity of a fourth type of potassium channel (the S channel), which is implicated in a form of learning in the sea slug *Aplysia*. Although various types of potassium channels are frequently the targets for neuromodulation, other channels are also influenced in this way. For example, DeRiemer *et al.* (*Nature* **313**, 313; 1985) recently showed that the action of a channel that controls calcium permeation is augmented when the enzyme protein kinase C is activated.

Why does the nervous system go to so much trouble to modify channel properties? One of the brain's chief powers is that it is self-modifying — one set of computing elements can, under neuronal control, be substituted for another while the basic nerve cells and their connections are maintained. Channel properties determine the information handling capabilities of neurones; altering them changes the way information is processed.

Even though we know that neuromodulatory phenomena are both common and important, we are only now starting to define the range of effects and the underlying mechanisms. In no case has the entire regulatory cascade been fully elucidated and much more must be done to work out enough specific examples before we can begin to perceive the underlying logic. Discovering how the brain modifies its own machinery is one of the most important challenges in modern neurobiology and is becoming one of the most active areas of research. □

Charles F. Stevens is in the Section of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

High-energy astrophysics

Is Cygnus X-3 a source of gamma rays or of new particles?

from Alan Watson

IF NEW results from the Soudan proton-decay project¹ are being interpreted correctly, cosmic, rather than man-made, accelerators are providing evidence either for a new elementary particle or for unexpected interactions of existing ones. The cosmic accelerator in question is the well-known X-ray and radio source, Cygnus X-3. By observing gamma rays above 10^{15} eV, groups at Kiel² and Leeds³ universities have identified Cygnus X-3 as one of the most powerful emitters of charged cosmic-ray particles in the Galaxy^{4,5}. These groups have shown that a few high-energy air-shower events from the direction of Cygnus X-3 display the distinctive 4.8-h modulation seen at X-ray and TeV gamma-ray energies. Because of the continuity of the spectrum from 10^{11} to 10^{15} eV (refs 2,3), and in the absence of other straight-

forward explanations, it has been generally assumed that these signals are gamma rays resulting from π^0 decay in the environs of the source. The new results from proton-decay experiments in the Soudan mine¹ and in the Mt Blanc tunnel (NUSEX)⁶ challenge this simple picture and may demand a change in our views about particle physics.

Cygnus X-3, one of the most luminous X-ray sources in the Galaxy, is a stellar-binary system some 12 kpc from the Sun⁷ in which a neutron star orbits a more massive companion star. X-rays are emitted when gas flowing from the massive star falls into the deep gravitational potential well surrounding the neutron star. The X-rays are modulated with a 4.8-h period which slowly lengthens and which is believed to be the orbital period of the system

— the minimum of the X-ray signal defines zero phase, $\phi = 0$, and is thought to occur when the neutron star is eclipsed by its companion. Signals, interpreted as gamma rays, have been seen regularly in many independent experiments from 10^{11} to 10^{16} eV, as discussed previously in these columns⁸. Here, the pattern of emission is variable but relatively sharp spikes of emission (of width $\Delta\phi \leq 0.1$) have usually been seen near $\phi = 0.2$ and/or $\phi = 0.65$. Nuclei accelerated in the electromagnetic fields of the neutron star may generate π^0 s, and hence gamma rays, through interactions in the circumstellar material⁵. Above 10^{11} eV, cosmic-ray primaries are detected through the cascades that they produce in the atmosphere. It is an inference to say that gamma rays have produced the showers, but neither charged particles nor neutrons could be the primaries — the interstellar magnetic field is much too turbulent for the former to retain their original direction or phase coherence, whereas neutrons of 10^{15} eV have a mean path length before decay of only 30 pc. It has long been expected that gamma-ray initiated air showers should be deficient in muons by a factor of about 10 compared with proton showers, because the photopion production cross-sections are 100 times smaller than proton-proton cross-sections. In an attempt to secure the gamma-ray identification, the Kiel group⁹ searched for signals from muons above 2 GeV in a shielded 65 m² detector within their array. It came as a shock that for the 14 'on-source' events for which they had data, the muon content was about 80 per cent rather than about 10 per cent of that in 'off-source' events: were the primaries really gamma rays? This difficulty was recognized but tended to be dismissed because the Kiel detector could not follow muon trajectories.

In this context the recent work by the Soudan group¹ demands attention. It is the first published report of a remarkable effect, apparently duplicated in the NUSEX experiment⁶, which has hitherto been known only to readers of the *samizdat* literature of particle physics. The Soudan and Mt Blanc groups operate large volume proton-decay detectors at depths of 1,800 and 5,000 metres water equivalent, respectively. Both find that those straight tracks which pass through their detectors (interpreted as single muons) and which project back close to the Cygnus X-3 direction, exhibit its distinctive 4.8-h periodicity. In the Soudan experiment the enhancement of 600 GeV muons is in the phase range $0.65 < \phi < 0.90$ and consists of 84 ± 20 events, whereas for NUSEX the excess of 1 TeV muons comprises 19 ± 5.7 events in the interval $0.7 < \phi < 0.8$. The possibility of chance giving the observed distribution in each experiment is estimated separately to be between 10^{-3} and 10^{-4} .

If Cygnus X-3 is a gamma-ray source, with the rays originating from π^0 decay, then it must also be a neutrino source; both the Soudan and Mt Blanc groups have con-

sidered the possibility of neutrinos as muon parents. The Soudan group find that the local zenith angle distribution of events in the phase peak is similar to that of ordinary muons and claim that they can reject completely the hypothesis of an isotropic zenith angle distribution expected for neutrino primaries. In the NUSEX experiment, the predicted neutrino-induced muon flux is too low by a factor of about 800 to explain the observations. To explain the Soudan results, Marshak *et al.*¹ propose either a new interaction or a new neutral particle. To maintain phase coherence, such a particle must have travelled from Cygnus X-3 with a Lorentz factor above 2×10^4 .

But some worries about the new experiments remain. The Soudan effect is maximized in a direction displaced by about 3° from the known radio position of Cygnus X-3, which is rather more than can be accounted for by uncertainties in the horizontal orientation of the detector. It is not clear that the background calculation correctly accounts for the possibility of variable exposure at different phases throughout the experiment. In particular the background summed over all phase bins is not equal to the source count as it should be for the background calculation described (J. Lloyd-Evans, personal communication). In the NUSEX analysis a $10^\circ \times 10^\circ$ area centred on the source has been required to maximize the signal; this is much larger than the $1^\circ \times 0.5^\circ$ error in celestial

coordinates estimated for each particle track. The absence of any signal above 7 TeV also needs explaining. In neither experiment is it clear whether any statistical degrees of freedom remain unaccounted for in the probability estimates.

As a final twist to this complex story the Akeno air-shower group has just announced a result¹⁰ agreeing with conventional views of the muon content of gamma-ray showers. In air showers with less than a thirtieth of the normal number of muons they find a 4.8-h modulated signal from Cygnus X-3 above 10^{15} eV and at phase 0.5–0.6. Both the phase and intensity agree with recent work from Haverah Park (A. Lambert *et al.*, manuscript in preparation), where confirmatory evidence will be sought with the University of Nottingham's new 40 m² muon detector. At the moment the gamma-ray interpretation of Cygnus X-3 is the simplest hypothesis and it may be premature to reject it. $\square$

1. Marshak, M.L. *et al. Phys. Rev. Lett.* **54**, 2079 (1985).
2. Samorski, M. & Stamm, W. *Astrophys. J.* **268**, L17 (1983).
3. Lloyd-Evans, J. *et al. Nature* **305**, 784 (1983).
4. Wdowczyk, J. & Wolfendale, A.W. *Nature* **305**, 609 (1983).
5. Hillas, A.M. *Nature* **312**, 50 (1984).
6. Battisoni, G. *et al. Phys. Lett. B* (submitted).
7. Dickey, J.M. *Astrophys. J.* **273**, L71 (1983).
8. Porter, N.A. *Nature News and Views* **305**, 179 (1983).
9. Samorski, M. & Stamm, W. *Proc. 18th Int. Conf. Cosmic Rays* **11**, 244 (1983).
10. Kifune, T. *et al. Astrophys. J. Lett.* (submitted).

Alan Watson is in the Physics Department, University of Leeds, Leeds LS2 9JT, UK.

Plant metabolism

Coping with drought by the sea

from Peter D. Moore

CRASSULACEAN acid metabolism (CAM) is a photosynthetic strategy characteristically associated with plants subjected to periodic or constant water stress. It is therefore, more usually described in plants of arid lands and with continental distributions than in those found in oceanic areas. It is particularly surprising that a recent study by P.P. Daniel *et al.*¹ of *Umbilicus rupestris* (wall pennywort) — a distinctly oceanic member of the Crassulaceae — should have revealed that this plant exhibits a physiological response to drought that is reminiscent of CAM.

During the past century it was observed that some succulent species of plants accumulate acid in their tissues during the night, but it was not until the 1940s that the full significance of this behaviour was understood. Leaf acidification is accompanied by the nocturnal uptake of carbon dioxide and it has become clear that there is a temporary fixation of carbon by the enzyme phosphoenolpyruvate carboxylase into a 4-carbon compound, oxaloacetate, which is subsequently converted to malate. In the daytime, the malate is decarboxylated and the released CO₂ is refixed by the Calvin cycle enzymes. The main advan-

tage of the pathway is the opportunity it provides for water conservation, because the stomata in the leaf need only open for gaseous exchange at night when transpiration losses can be kept to a minimum.

The efficiency of water conservation in CAM plants becomes clear when their transpiration/photosynthesis ratios (grams of water transpired per gram CO₂ fixed) are compared with those from plants that have other photosynthetic strategies². For CAM plants the ratio is usually 50–100, for C₄ plants it is 250–300 and for C₃ plants 400–500. The strategy is also efficient at conserving carbon, as the CO₂ generated by respiration can be trapped and retained in the same manner.

One of the costs of efficient water conservation in CAM plants is a fairly slow rate of carbon accumulation in the growth of tissues. Therefore, there are clear energetic advantages in the adoption of a flexible system in which it is possible to switch to CAM only during periods when it is of advantage to the survival of the plant, such as in times of water stress; several species of plant have now been described that can do this³. One is the succulent *Mesembryanthemum crystallinum*⁴, found on

sandy shores and sand dunes around the Mediterranean. Other species switch to an incomplete form of CAM when they experience drought. This has been termed 'CAM idling'. In this mode, there is no fixation of atmospheric carbon during the night, but there is a nocturnal accumulation of acid resulting from the fixation of respiratory CO_2 . Here, carbon economy is closely linked to water conservation.

It is this type of CAM idling that has now been observed in *Umbilicus rupestris* when it is exposed to drought. Under normal well-watered conditions, *Umbilicus* behaves like a conventional C_3 plant as far as its carbon exchange is concerned, taking up atmospheric CO_2 during the day and showing a net output of CO_2 at night. After six days of drought, the daytime up-take becomes confined to a short period in the morning, after which stomatal exchange is curtailed. The plant makes up the reduced carbon intake by lowering the respiratory losses at night when phosphoenolpyruvate carboxylase traps the CO_2 generated in respiration and converts it to oxaloacetate and then malate, resulting in tissue acidification.

Evidently the falling water potentials in the plant activate a metabolic switch, but the nature of the operation of this switch is still unknown. One of the consequences of switching is the enhancement of carboxylase enzyme activity; Daniel *et al.* have shown that the properties of the activated enzyme are somewhat modified — for example, it is less sensitive to low pH and to malate inhibition⁵. But they do not suggest the generation of a new separable form of the enzyme.

There remains the intriguing ecological question of why an oceanic species should possess this unusual capacity for dealing with drought. The answer must lie in its preferred habitat, for *Umbilicus rupestris* is a plant of cliff crevices and stone walls,

Wall pennywort growing epiphytically on a date palm in Seville, southern Spain.



and sometimes even grows as an epiphyte on larger plants. It can be found at the southern extremities of its range in southern Spain growing on the coarse trunks of the date palm, *Phoenix dactylifera* (see figure). In such exposed locations, often with little or no true soil around its roots, the risk of drought must be very real for *Umbilicus*, even in the wetter regions of north-west Europe. Clearly, the capacity to switch to CAM idling enables the plant to make the most of any given set of conditions. Discovery of CAM in an oceanic temperate plant forms an interesting equivalent to the tropical epiphytic ferns in which CAM has previously been described⁶. □

1. Daniel, P.P., Woodward, F.I., Bryant, J.A. & Etherington, J.R. *Ann. Bot.* **55**, 217 (1985).
2. Edwards, G. & Walker, D.A. *C₃, C₄ Mechanisms and Cellular and Environmental Regulation of Photosynthesis*. (Blackwell Scientific, Oxford, 1983).
3. Ting, I.P. & Rayder, L. in *Crassulacean Acid Metabolism* (eds Ting, I.P. & Gibbs, M.) 193 (American Society of Plant Physiologists, Rockville, 1982).
4. Winter, K., Lüttge, U., Winter, E. & Troughton, J.M. *Oecologia* **34**, 225 (1978).
5. Daniel, P.P., Bryant, J.A. & Woodward, F.I. *Biochem. J.* **218**, 387 (1984).
6. Wong, S.C. & Hew, C.S. *Am. Fern J.* **66**, 121 (1976).

Peter D. Moore is in the Department of Plant Sciences, King's College, 68 Half Moon Lane, London SE24 9JF, UK.

Particle physics

Look before you LEP

from David J. Miller

THE physics agenda for the large electron-positron machine (LEP), now being constructed at the European Organization for Nuclear Research (CERN), was first circulated in the late 1970s. At that time the Z^0 and $W^\pm$ were entirely theoretical particles, but they were needed by the electroweak theory of Glashow, Salam and Weinberg to carry the weak part of the force. Now that these particles have both been discovered by experimenters using the proton-antiproton ($p\bar{p}$) collider at CERN, and after the other developments of the past six years, what should the goals of LEP be? John Ellis convened a Physics Jamboree at CERN on 19 March 1985 to examine this question. Similar answers came from the five working-groups that

reported: we must test the validity of the standard model theories at energies where they might begin to break down, and we must probe for the explanation of their many arbitrary parameters.

The important feature of an electron-positron (e^+e^-) machine like LEP, compared with the existing $p\bar{p}$ collider, is that there are no spectator particles to complicate the main event. Although Z^0 and $W^\pm$ particles can be found in $p\bar{p}$ collisions, thorough investigations of their properties will only be possible with an e^+e^- machine. Guido Altarelli (Rome) reviewed the precision studies that can be made when the e^+e^- energy is around 94 GeV, equal to or just above the Z^0 mass. In this energy region nearly all reactions will go

via a resonance centred at the Z^0 particle and LEP becomes a ' Z^0 factory'. Every possible pair of particle and antiparticle into which the Z^0 can decay will be produced at roughly comparable rates — for example, the heavy tau leptons, the rare and heavy beauty quarks or the everyday down quarks. Although the rarer quarks and leptons are much heavier than the common ones, their partner neutrinos all seem to be very light, so the Z^0 could decay to particle-antiparticle pairs of new neutrinos that are the partners of still undiscovered quarks and leptons. This should allow us to prove the number of generations of heavier and heavier sets of quarks and leptons that still remain to be found after the 'beauty-top-tau' generation. Altarelli reported a technique which the LEP experimental teams believe they can use to make a definitive measurement of the number of neutrino types, and hence the number of generations, since there seems to be one neutrino per generation. This would require only ten days of experimental running when LEP has achieved its planned beam intensity.

Although the Stanford Linear Collider (SLC) in California, now also under construction, may be the first e^+e^- machine to produce the Z^0 , its initial detector will probably not be good enough to make this particular measurement. Some of the other physics around the Z^0 mass will probably be investigated first by SLC groups, but it will require very patient and precise work to check all the predictions of the electroweak model — particularly to measure the higher-order corrections to its fundamental constant, the Weinberg angle. Here, the cross-checks between SLC and the four LEP experiments will be important.

Wilfried Buchmüller (CERN) described the feasts in store when we finally pass the toponium threshold — the energy at which toponium, a bound state of a top quark with an anti-top quark, can be created. Last year the UA1 experiment at the $p\bar{p}$ collider gave an estimated mass for the top quark of between 30 and 50 GeV, but no independent confirmation has yet appeared for this rather imprecise result. Toponium may therefore be found anywhere from the present limit of 43 GeV, set by the experiments at the PETRA e^+e^- collider in Hamburg, to the highest possible LEP energy. It should have many of the properties of the two well-established positronium-like quark-antiquark states, the J/psi (charm-anticharm) and the upsilon (beauty-antibeauty). Although toponium should have more excited states than the J/psi and the upsilon, all but the lowest of them will be impossible to resolve, even if they lie well below the Z^0 mass. If they lie close to the Z^0 mass, the physics could be extremely complicated since the Z^0 and toponium states would behave like nearby resonances coupled to the same input channel. They would have opposite phases, so toponium would appear not as localized bumps in the cross section but as narrow dips on the

shoulder of the larger Z^0 bump. Because of the greater mass of the top quark, toponium will be a smaller system than either J/ψ or ψ , and its detailed level-structure will give evidence for the behaviour of the quantum-chromodynamic potential at very short range. Other ways of investigating quantum chromodynamics in e^+e^- annihilation and the scattering of virtual photons from virtual photons, were reviewed by Ahmed Ali (DESY, Hamburg).

The totally new effects that might appear once the natural energy scale of the weak interactions is reached were reviewed by John Ellis (CERN). The cause of the large masses of the $W^\pm$ and Z^0 , together with the generations of heavy quarks, is one of the great open questions in current theory. These masses may be generated by the Higgs-Kibble mechanism, which introduces a new heavy scalar particle that pervades space, modifying the vacuum and coupling most strongly to the heavy particles. No one has yet identified corresponding free Higgs bosons, although many strategies to look for them have been suggested, especially in Z^0 or toponium decay. But other more complicated theoretical schemes exist which would not necessarily need simple Higgs particles. If their thresholds are reached at LEP, the effects may be seen in anomalous events like the UA1 mono-jets, which some have tried to explain in terms of supersymmetric theories (see J. Ellis, *Nature* 313, 626; 1985).

Herwig Schopper, the director-general of CERN, announced that he plans an early upgrading of LEP from the initial (50 + 50) GeV LEP 1 machine, which starts work in 1989, to the (100 + 100) GeV LEP 2. The only change needed would be to replace the accelerating radiofrequency cavities. Great progress has been made in developing the required superconducting cavities and some of them will even be included in LEP 1.

Schopper's announcement was particularly interesting after the discussion of the physics to be done at LEP 2 energies (Roberto Peccei, DESY). The bankable study, which cannot be done on any other machine, is to produce W^+W^- pairs directly via the fundamental couplings of electroweak theory. This gives yet another probe of higher order corrections. If the Higgs particle has not yet been seen, it will only be through such higher order electroweak studies that we will be able to differentiate between the different candidate models that generate masses. There is also more speculative work to be done at the higher LEP 2 energies, continuing the search for the effects of supersymmetry or for some new heavier scale of composite substructure that may underlie the present particles. The Jamboree showed that LEP has an even longer list than before of important specific questions to answer. If there are surprises as well, it will be a bonus. □

David J. Miller is in the Department of Physics and Astronomy, University College London WC1 6BT, UK.

Ocean Drilling Program

Shakedown and sea trials

from the Leg 100 shipboard scientific party*

THE Ocean Drilling Program (ODP), successor to the Deep Sea Drilling Project (DSDP), commenced its field operations on 11 January 1985 with an 18-day shakedown and sea-trials cruise (Leg 100), departing from Pascagoula, Mississippi and arriving at Miami, Florida. The cruise came after an extensive conversion and overhaul of the newly chartered drilling vessel *Joides Resolution*. This drillship is 470 feet long and 70 feet wide with a displacement of 16,596 tons, a derrick towering 217 feet above the waterline, and the ability to employ a drillstring of 30,000 feet (9,146 m) in water depths up to 27,000 feet (8,632 m).

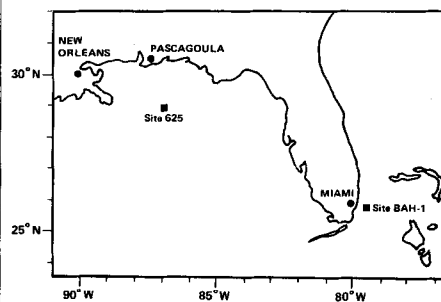
Prior to the shakedown cruise, a seven-storey main laboratory structure, geophysics laboratory and a separate library and study area were constructed aboard the vessel. Additional scientific, office and storage space has been provided; major modifications to the ship's dynamic positioning system, which enables the vessel to maintain a relatively fixed position over the hole being drilled in the sea bottom, now enable long-based and short-based line systems, as well as the original ultra-short-based line system, to be used. The system is supported by 12 powerful 750 horsepower thrusters, as well as by two main shafts each with six 750 horsepower motors. In addition, a 400-ton heave compensator, which keeps the drillstring stable relative to the seafloor, has been installed. These are only a few of the numerous modifications made to the original rig.

The science laboratories contain a complete array of seagoing research equipment that will allow shipboard scientists to carry out a wide range of physical and chemical analyses on the core samples recovered by drilling. There are laboratories to study sedimentology, physical properties, palaeomagnetism, chemistry, gas analysis, palaeontology, petrology, thin section, X-ray diffraction and fluorescence microscopy, scanning electron microscopy, and downhole measurements. In addition, facilities for photography, electronics, and refrigerated core storage are available. While the ship is in transit between drill sites, digital single-channel seismics are recorded and processed in the on-board geophysics laboratory. There is also a research-oriented computer system which is designed to perform as many routine clerical and arithmetical tasks as possible. A pair of VAX computers serve as a central processor and data library for 50 microcomputers distributed throughout the laboratories.

During the inaugural cruise, three holes were drilled at site 625, near De Soto Canyon, west Florida Shelf (see figure). The scientific objectives were to document

the sedimentological, palaeontological, geochemical, geotechnical and geomagnetic characteristics of the sedimentary sequences, and to correlate alternating sedimentary and erosive sequences to worldwide changes in sea level over the past several million years. The principal operational objectives were to test the rotary advanced piston coring (APC), and extended core barrel (XCB) drilling and coring systems, and to familiarize the drillship's crew with core handling and sampling procedures.

The deepest of the three holes drilled was 625B, which penetrated to 235.2 m sub-bottom. This hole was continuously cored with the APC system through 197.1 m of Plio-Pleistocene section. The XCB system continued to termination depth in the Lower Pliocene. At the third hole (625C), an attempt was made to obtain a complete section of the uppermost Quaternary by overlapping the HPC cores taken in the previous hole between 5 and 44.5 m sub-



Location map for site 625.

bottom. A visual comparison between the overlapping piston cores indicates that the sequences in each are closely matched. On the basis of physical properties and magnetic susceptibility data, we believe that no more than about 10 cm of material remained unsampled after the double HPC coring. Since this operation was accomplished in hostile sea conditions, it clearly demonstrates the usefulness and importance of the heave compensator.

During Leg 100, the onboard scientific, drilling and operational equipment was tested under varying sea conditions and generally performed to, or even exceeded, expectations. A fairly large number of adjustments or rearrangements of equipment were required but the minor nature of these is illustrated by the fact that the ship was able to depart on 31 January, after only two days in port, for the first fully operational, internationally staffed cruise of the Ocean Drilling Program, Leg 101. □

*P.D. Rabinowitz, W.J. Merrell, L. Garrison, R. Kidd, A. Adamson, C. Auroux, J. Baldauf, B. Clement, R. Merrill, A. Meyer, A. Palmer and E. Taylor, Ocean Drilling Program, Texas A&M University, College Station, Texas 77843, USA.

Split-gene evolution

Exon shuffling and intron insertion in serine protease genes

from John Rogers

EVER since the discovery of split genes, there has been a debate about why they are split. This can be resolved into three separate problems: the origin of the introns that split the genes (separating exons from each other), the role of introns in evolution, and their present function, if any (for example, refs 1–3). Ideas on these matters can be tested against the serine proteases (proteins that have particularly clear patterns of domains^{4,5}) now that more than a dozen of their genes have been sequenced. In agreement with one hypothesis, the different domains are encoded in separate exons, some of them apparently transferred from unrelated genes. But it is also clear that new introns have been inserted at various times in evolution.

Discussions of the role of introns in evolution have centred round ideas of 'exon shuffling', in which exons are supposed to be units in the evolutionary rearrangements of genes^{1–3}. One should bear in mind that exon shuffling, however advantageous in the long run of evolution, cannot explain the origin of introns, nor does it amount to a present-day function for them. Nevertheless, there has been

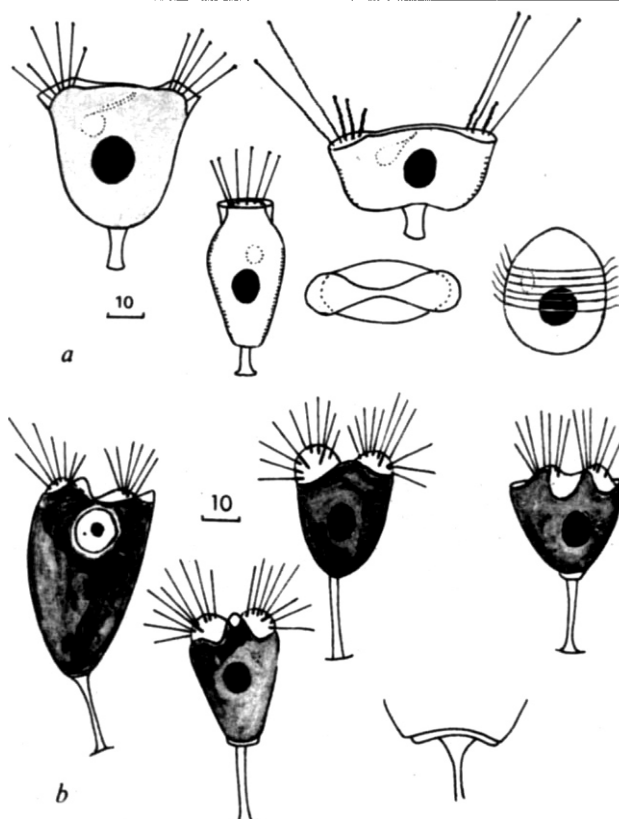
considerable interest in the idea that some exons encode protein domains, or at least coherent structural motifs, which would be good material for evolution by exon shuffling. Thus, they could be subject to duplications within a gene, as has happened in several genes such as that for fibronectin^{6,7} (and the serine proteases — see below). In addition, exons could be transferred between unrelated genes to create new proteins out of old domains. Attempts to validate this more interesting possibility have not been successful. For example, the nucleotide-binding fold of metabolic enzymes is not encoded separately^{2,8}, and the amino-terminal signal peptide of secreted proteins, which is often more or less contained in a single exon, is too variable in sequence for its ancestry to be reliably traced. But good evidence of exon shuffling is now available from the exon/intron organization of the serine protease genes.

The serine protease superfamily^{4,5} includes digestive enzymes (such as trypsin), enzymes of the blood-clotting and complement cascades and various other polypeptide-processing enzymes such as the

kallikreins. (To escape from the semantic jungle of 'pre', 'pro' and 'gen' terms, I use the names of the active forms of these molecules.) New serine proteases have been arising by gene duplications for most of biological history. Those just mentioned probably diverged more than 500 million years ago, and there are even more divergent homologues in bacteria^{5,9}. The protease domain itself is activated by cleavage from a larger precursor. The non-enzymatic portion often includes various structurally distinct domains with particular physiological functions.

Exon shuffling is best illustrated in the 5' halves of the dozen sequenced genes, as shown in the figure. In each case, the coding region outside the protease domain is split by at least one intron in phase I of the reading frame (see figure) and the various extra domains within the region are encoded by separate exons or groups of exons, also separated by introns in phase I. It is presumably this coincidence of phasing that has permitted the assembly of diverse exons and the tandem duplications of some of them. An example is the kringle domain (so named because of its topological resemblance to a traditional type of Scandinavian cake). The region of the gene that encodes the kringle, flanked by phase I introns, has been duplicated independently in the genes for thrombin¹⁰, tissue plasminogen activator¹¹ and plasmin¹² (not shown). This has happened even though the domain has different internal splicing patterns in these three genes^{10–12} (intron insertion?), and even though the donor splice site following the kringle seems to have been replaced by a splice site that is nine codons further downstream in the gene for uPA, another serine protease¹³ (intron sliding).

The tPA gene contains another two exons which, as reported by Ny *et al.*¹¹, provide the first example of exon transfer between otherwise unrelated genes. The third coding exon, unique to tPA, encodes a disulphide-linked 'finger' that is homologous to the tandemly repeated 'fingers' of the fibronectin molecule¹⁴. In fibronectin, each 'finger' is encoded in a separate exon⁶, and although their boundaries have been mapped only approximately, the one example sequenced⁶ does have a potential splice site in a position precisely equivalent to that in tPA. Thus, the tPA exon seems to have come from a fibronectin gene. The fourth coding exon of tPA encodes a region of homology to epidermal growth factor (EGF)^{11,14}; a homologous exon is duplicated in the gene for clotting factor IX¹⁵. This exon may have come from a gene like the EGF gene, which contains several EGF-like segments in separate exons (G. Bell, personal communication); but any firm conclusion must await the sequencing of the EGF gene. Even further back, the fibronectin-related and EGF-related exons of all these genes may have had a common origin, as all three exons encode sequences approximating to Cys-



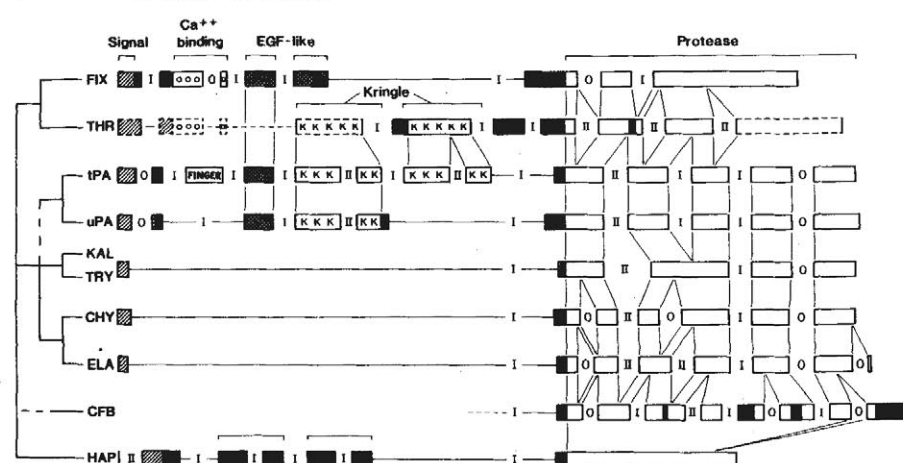
Two species of suctorian ciliate protozoa from a revision of the genus *Acineta* and its morphological relatives by C.R. Curds (*Bull. Br. Mus. nat. Hist. (Zool.)* 48, 75; 1985). *a* *Acineta gammari*, lateral, end and top views and ciliated larva (far right); *b* *Acineta dentata*. Bars are 10 μ m.

X-Cys-X-X-Gly_{Asn}-X-X-Gly-X-X-Cys near their 3' ends¹⁴.

The domain that encodes the protease activity itself sheds light on the origin of introns. As in several other superfamilies, the positions of introns are different in different genes, and the question is whether this results from insertions or excisions. Clearly introns can be excised, at least in mammals^{1,16,17}; that they can also be inserted, at least in invertebrates, is suggested by the apparently random positioning of introns in invertebrate gene families such as actin¹⁶ and myosin¹⁸, but this idea has never been proved. The different serine proteases, which are also pre-vertebrate in origin, show an unambiguous history of intron insertions (see figure).

This is clearest within the trypsin family (middle five lines in the figure), the family tree of which, as deduced from protein sequences, is shown at the left. These genes all have three introns in common. As one follows the branches of the family tree, new unique introns appear in the different genes — one extra in tPA and uPA, two extra in chymotrypsin and three extra in elastase. It seems that the shared introns were also inserted, as the introns in the trypsin family are entirely different from those in thrombin, and almost all are different from those in complement factor B. Such unrelated patterns could hardly have arisen by random deletions of introns. Therefore, most or all of the introns in the protease domain have been inserted. (The only apparent exceptions are the two introns of the factor IX protease domain, which match introns in elastase and factor B, respectively. But according to the amino-acid homologies, the first of these is in a segment that may have come from a chymotrypsin- or elastase-like gene by recombination, while the second is in a region of weak and uncertain relationships.)

Given that these introns have been inserted, do their positions show us anything about how or why this happened? Some regularities can be discerned. For one thing, the figure suggests that new insertions tend to occur near the middle of pre-existing exons. Thus a gene evolves towards a uniformity of exon size. This has been noted independently for the elastase gene¹⁹ and for genes in general^{8,20}. Another regularity^{3,21} is that many of the introns map to surface loops, and particularly to variable-length loops, in the protein structure. Craik *et al.*^{3,21,22} suggested ways in which this length variation could be caused by the presence of the introns, for example by substitution of alternative splice sites (intron sliding). Another possibility now worth considering is that introns generate length variation because they are inserted inaccurately. As introns do not resemble any known transposable elements, one can speculate that a new intron might be included as part of a longer insertion, or might be induced when an insertion activates a cryptic splice site that is nearby.



Structures of the coding parts of serine protease genes. The family tree as deduced from protein sequencing^{5,9,26} is summarized schematically at left, the dashed branches being tentative. FIX is human clotting factor IX¹⁵; THR, human thrombin¹⁰ (gene structure not reported for regions shown dashed); tPA, human tissue plasminogen activator¹¹; uPA, pig urokinase¹³; KAL, mouse kallikrein; TRY, rat trypsin (gene structures are identical for two trypsin genes²² and four kallikrein genes, including nerve growth factor subunits^{27,28}); CHY, rat chymotrypsin²⁹; ELA, rat elastase¹⁹ (two similar genes); CFB, human complement factor B³⁰ (the exons shown are preceded by other, non-homologous exons); HAP, human haptoglobin²³; Exons are shown as boxes shaded according to their homologies. Sequences with no discernible homology to each other are in black. Introns are not to scale, but their phase relative to the reading frame is indicated, according to whether they follow the first (I), second (II), or third (O) nucleotide of a codon.

On the other hand, some introns are in conserved coding regions, and these include some of the most striking cases of independent insertions into the same region. Thus, five genes in the figure contain an intron that is 14–21 codons into the protease domain, but at least three of these are in different positions relative to codons for conserved amino acids. Even more remarkable, the penultimate introns in complement factor B and elastase fall in the same conserved glycine codon, but one nucleotide apart. These coincidences cannot reasonably be explained by intron sliding; this would require vanishingly improbable coincidences of mutations in order to shift an intron within a conserved coding region, or to shift the phase of an intron. The non-random positions of these introns seem to imply either sequence specificity at the insertion, or selection following it, for reasons that are still mysterious. The methodological lesson, however, is clear. Correlation of introns with protein structure^{1,8} does not, *per se*, tell us whether the introns are original or recent additions.

A final puzzle is presented by the gene for haptoglobin²³, a member of the superfamily but not known to have protease activity⁹. Although this gene has plenty of introns in its 5' half, dividing duplicated phase I exons as in the other genes, there are no introns in the 3' half, which encodes the protease-like domain. This could represent the condition of the ancestral serine protease gene, intronless-like genes of bacteria. Alternatively, it could represent a much more recent processed gene, derived from the messenger RNA of a split serine protease gene via reverse transcription, which became attached to a different split gene. A similar history has been

postulated for part of the (unrelated) clotting factor VIII gene²⁴; this mechanism has also been proposed as a general way of evolving new proteins²⁵. Processed genes are common only in mammals¹⁷, so if one is incorporated either in haptoglobin or in factor VIII, the gene from which it was derived may still be present and detectable by nucleic acid hybridization. □

- Blake, C. *Nature News and Views* 306, 535 (1983).
- Cornish-Bowden, A. *Nature* 313, 434 (1985).
- Craik, C.S. *et al.* *Nature* 299, 180 (1982).
- Neurath, H. *Science* 224, 350 (1984).
- Young, C.L. *et al.* in *Atlas of Protein Sequence and Structure* (ed. Dayhoff, M.O.) Vol. 5, suppl. 3, 73 (1978).
- Hirano, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 80, 46 (1983).
- Kornblith, A.R., Vibe-Pedersen, K. & Baralle, F.E. *Nucleic Acids Res.* 12, 5853 (1984).
- Lonberg, N. & Gilbert, W. *Cell* 40, 81 (1985).
- Kurosky, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 77, 3388 (1980).
- Degen, S.J.F., MacGillivray, R.T.A. & Davie, E.W. *Biochemistry* 22, 2087 (1983).
- Ny, T., Elgh, F. & Lund, B. *Proc. natn. Acad. Sci. U.S.A.* 81, 5355 (1984).
- Malinowski, D.P., Sadler, J.E. & Davie, E.W. *Biochemistry* 23, 4243 (1984).
- Nagamine, Y. *et al.* *Nucleic Acids Res.* 12, 9525 (1984).
- Banyai, L., Varadi, A. & Pálffy, L. *FEBS Lett.* 163, 37 (1983).
- Anson, D.S. *et al.* *EMBO J.* 3, 1053 (1984).
- Fornwald, J.A. *et al.* *Nucleic Acids Res.* 10, 3861 (1982).
- Rogers, J.H. *Int. Rev. Cytol.* 93, 187 (1985).
- Karn, J., Dibb, N. & Miller, D. in *Cell & Muscle Motility* (ed. Shay, J.) 6, 185 (1984).
- Swift, G.H. *et al.* *J. biol. Chem.* 259, 14271 (1984).
- Naora, H. & Deacon, N.J. *Proc. natn. Acad. Sci. U.S.A.* 79, 6196 (1982).
- Craik, C.S., Rutter, W.J. & Fletterick, R. *Science* 220, 1125 (1983).
- Craik, C.S. *et al.* *J. biol. Chem.* 259, 14255 (1984).
- Maeda, N. *et al.* *Nature* 309, 131 (1984).
- Gitschier, J. *et al.* *Nature* 312, 326 (1984).
- Brenner, S. in *The Evolution of DNA Sequences* (eds Clarke, B.C. *et al.*) (in the press).
- Katayama, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 76, 4990 (1979).
- Mason, A.J. *et al.* *Nature* 303, 300 (1983).
- Evans, B.A. & Richards, R.I. *EMBO J.* 4, 133 (1985).
- Bell, G.I. *et al.* *J. biol. Chem.* 259, 14265 (1984).
- Campbell, R.D. & Porter, R.R. *Proc. natn. Acad. Sci. U.S.A.* 80, 4464 (1983).

John Rogers is at the Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Igneous petrology

Discordance in layered intrusions

from R.S.J. Sparks

LAYERED intrusions are the frozen remains of magma chambers. Their distinguishing characteristic is a vertical variation in mineral abundance, giving rise to a horizontal layering superficially resembling that seen in sedimentary rocks. The study of their geology and petrology has played a central role in the development of ideas on the origin of magmas. Intrusions such as the Skaergaard in Greenland provide almost perfect natural examples of crystallization under closed-system conditions. Other intrusions, such as Rhum in Scotland and Stillwater in Montana, record the history of long-lived open-system magma chambers. The classic method of investigating a layered intrusion is to collect samples through the stratigraphy and to document variations in the composition and abundance of minerals. Usually one vertical section has been taken to represent the entire intrusion. In a recent paper, J. R. Wilson and S. B. Larsen¹ describe the first comprehensive two-dimensional study of a layered intrusion — the Fongen-Hyllingen intrusion in Norway. The relationships they report provide important new facts which any theory on the origin of layered intrusions must seek to explain.

Detailed geological mapping and systematic petrology provide the key to understanding large layered intrusions, which typically cover areas of hundreds of square kilometres. This kind of study is thus a substantial and time-consuming enterprise. Indeed a layered intrusion often becomes associated with one particular scientist who may have devoted considerable time to its study; perhaps the most notable link is between L.R. Wager and the Skaergaard intrusion. Wilson and colleagues at Aarhus University in Denmark have devoted over fifteen years to investigating the Fongen-Hyllingen intrusion and they have now produced one of the most significant studies of a layered intrusion for many years.

The variation in mineral composition observed in these intrusions is termed cryptic layering, whereas the presence or absence of minerals defines the phase layering. On the scale of outcrops, the layering is made visible by small changes in the relative abundances of minerals, here termed modal layering. Wilson and Larsen document dramatic lateral variations in mineral composition and show that the modal layering observed in the field is strongly discordant to cryptic and phase layering.

The Fongen-Hyllingen intrusion was emplaced as a sill-like body at 12–20 km depth during the Caledonian orogeny, roughly 425 Myr ago. The intrusion and its metamorphosed envelope of basic and pelitic country rocks are now part of a nappe complex and prolonged erosion has

exposed a longitudinal cross-section 33 km in length. The northern part (the Fongen and Ruten Series) has the greatest stratigraphical thickness (6.2 km) and is complicated by folding. Primitive ultramafic rocks occur at the base of the sequence and there are many reversals in the mineral compositions, a certain sign of an open-system magma chamber. The southern part (the Hyllingen Series) consists of 3.6 km of modally layered rocks, and is documented in detail by Wilson and Larsen.

Both the floor (the western margin) and the roof (the eastern margin) are exposed in the Hyllingen area. Geological mapping reveals intimate interfingering of country rock with layered rocks in the southern region. Swarms of huge country-rock inclusions, with individual metabasic xenoliths up to 1 km long and a few hundred metres wide, are found throughout the sequence. The layering is completely undisturbed by these xenoliths and they seem to be in their original positions rather than having fallen in from the roof or walls. The igneous rocks are highly evolved ferro-gabbros and ferrodiorites, with quartz-bearing syenites occurring next to the roof. Minerals include olivine, Ca-rich and Ca-poor pyroxenes, amphibole, plagioclase, Fe-Ti oxides, zircon, apatite, biotite and allanite. The very high Fe/Mg ratio of ferromagnesian minerals shows that all the rocks have crystallized from highly differentiated magmas. In any one vertical profile four stages of magma evolution can be recognized. Stage I is an unlayered gabbro. Stages II and IV record initial open-system behaviour with relatively constant mineral compositions followed by progressive differentiation. Stage III in the middle of the sequence records a major reversal with minerals gradually becoming less fractionated towards the top of the section.

The most important results of the study are the recognition of lateral variations in mineral compositions, which are almost as great as those recorded in vertical profiles and provide evidence by changes of mineral abundances on millimetre or centimetre scales. These layers strike north-south, parallel to the floor and roof with no deviation at the southern contact or locally around enormous xenoliths. However, mineral compositions (defining cryptic layering) change dramatically along the strike with minerals becoming more evolved towards the southern margin. For example, at the boundary between stages III and IV the olivine and plagioclase vary from Fo₇₅: An₆₃ to Fo₁₃: An₄₂ over a 10 km strike length. The stage boundaries are parallel to field layering, showing that changes in the differentiation trend occurred simultaneously at the same

stratigraphical level. With eleven separate vertical profiles Wilson and Larsen have been able to contour lines of constant mineral composition. The contours are discordant to the modal layering by angles which are typically 20° but are as much as 90° at the southern margin. As described in their preliminary account², boundaries recording the appearance of new cumulus minerals (zircon and apatite) are also highly discordant to the rhythmical layering.

Conventionally, the visible modal layering is thought to represent the trace of the ancient magma-chamber floor and therefore each layer is a time-plane. If this is so then the implication of the discordance is that the magma chamber must have been compositionally stratified throughout its evolution. Wilson and Larsen argue that the dramatic lateral gradients and the lack of disturbance of layering by xenoliths eliminate crystal settling. They believe that the crystals grew *in situ* along the floor. They propose that the floor was gently tilted so that increasingly differentiated magma in a zoned chamber was in contact with the floor in the (southerly) upslope direction. Thus, the minerals systematically change composition and new phases appear up the slope. Addition of magma from the north complicated the picture by elevating the zoned magma column and causing gradual reversals as lower and more primitive magma migrates up the slope. A more elaborate model, based on the work of Irvine, Keith and Todd³, envisages lateral accretion of layers, in which the field layering, mineral-composition layering and the accretion front (the time-plane) are all discordant.

Recent petrological literature contains several new ideas on the processes that form layered intrusions. Several workers have argued that double-diffusive convection plays an important role^{3–6}. There is also field and theoretical evidence that compaction of cumulate piles and post-cumulus migration of melt^{7,8} can lead to changes in mineral compositions, which further complicates interpretations of their petrogenesis. Irrespective of the merits of recent thinking and speculation, the discoveries of major lateral variations and discordances in layered intrusions must be central facts on which different hypotheses will succeed or founder. Wilson and Larsen's work should herald a new phase of research in layered intrusions, in which documentation of lateral changes will be as important as vertical variations. □

1. Wilson, J.R. & Larsen, S.B. *Geol. Mag.* **122**, 97 (1985).
2. Wilson, J.R. & Larsen, S.B. *Nature* **299**, 625 (1982).
3. Irvine, T.N., Keith, D.W. & Todd, S.G. *Econ. Geol.* **78**, 1287 (1983).
4. McBirney, A.R. & Noyes, R.M. *J. Petrol.* **20**, 487 (1979).
5. Huppert, H.E. & Sparks, R.S.J. *A. Rev. Earth planet. Sci.* **12**, 11 (1984).
6. Turner, J.S. & Gustafson, L.B. *Econ. Geol.* **73**, 1082 (1978).
7. Irvine, T.N. in *Physics of Magmatic Processes* (ed. R.B. Hargreaves) 325 (Princeton University Press, 1980).
8. McKenzie, D.P. *J. Petrol.* **25**, 713 (1984).

R.S.J. Sparks is in the Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK.

Has inductive probability been proved impossible?

SIR—The interesting letter by Popper and Miller¹ and the subsequent correspondence²⁻⁴ leaves in some doubt the question whether a proof of the impossibility of inductive probability been given. As this would be a remarkable achievement, it is no criticism of these authors that we raise this question again. In our view the answer is a clear "No". We shall follow your other correspondents in not burdening the discussion with references to extensive earlier literature.

The key elements in the Popper-Miller (PM) argument are: first the evidence e , when found, counter-supports the implication $h \leftarrow e$ of a hypothesis h by the evidence; and second, this is due to the negative nature of that part of the support for $h \leftarrow e$ arising from e which is not purely deductive. From this type of consideration, PM infer the impossibility of inductive probability. In both the original PM and the following discussion, the part played by the negation $\bar{e}$ was not considered explicitly. Yet it is helpful to do so in showing what is going on in these arguments, as will now be explained. Two basic results, to be formulated as "theorems", are useful preludes to this explanation.

Theorem A

$$\begin{aligned} p(h \leftarrow e|e) &\leq p(h \leftarrow e|e) + p(\bar{h}|e)p(\bar{e}) \\ &= p(h \leftarrow e) \leq p(h \leftarrow e|\bar{e}) = 1 \\ p(h \vee e|\bar{e}) &\leq p(h \vee e|\bar{e}) + p(\bar{h}|\bar{e})p(e) \\ &= p(h \vee e) \leq p(h \vee e|e) = 1 \end{aligned} \quad (1)$$

Theorem B The following have the same sign:

$$\begin{aligned} p(h|e) - p(h) &= p(h|e)p(\bar{e}) - p(h|\bar{e})p(\bar{e}) \\ &= p(\bar{h}|\bar{e})p(\bar{e}) - p(\bar{h}|e)p(\bar{e}) \\ &= p(\bar{h}) - p(\bar{h}|e) \end{aligned} \quad (2)$$

and

$$\begin{aligned} p(\bar{h}|\bar{e}) - p(\bar{h}) &= p(\bar{h}|\bar{e})p(e) - p(\bar{h}|e)p(e) \\ &= p(h|e)p(e) - p(h|\bar{e})p(e) \\ &= p(h) - p(h|\bar{e}). \end{aligned} \quad (3)$$

Theorem A shows the origin of theorem 1 of PM that

$$p(h \leftarrow e) > p(h \leftarrow e|e) \quad (\text{if } p(e) \neq 1) \quad (4)$$

as follows. The underlined terms are positive, so that (4) could be incorporated in (1). Note that $p(h \leftarrow e) = p(h \vee \bar{e})$ allows for the possibility of $\bar{e}$, which possibility is ruled out by the assumption that e has occurred. This makes the inequality (4) intuitive as a logical necessity (which is incidentally independent of any inferential relation between e and h).

Since $h \leftarrow e \equiv h \vee \bar{e}$, $\bar{e}$ must "support" $h \leftarrow e$ maximally, or $p(h \leftarrow e|\bar{e}) = 1$. It is

intuitively to be expected, therefore, that e "counter supports" $h \leftarrow e$. These statements interpret Theorem A (an extension of theorems 1 and 2 of PM).

Theorem B shows relations between four "support" functions whose definition

$$s(h, e) = p(h|e) - p(h) \quad (5)$$

was introduced into the correspondence by Jeffrey³. They are shown as differences in different ways. Adopting the nomenclature of your correspondents (which is possibly misleading), we call the underlined terms "non-deductive". The introduction of negations is again helpful here, for they show, using (5), that the "non-deductive" part in (2) appears with a positive sign (support) in $s(h, e)$, but with a negative sign (counter support) in $s(\bar{h}, e)$. Similarly for (3). Thus the remark by PM that all probabilistic support is purely deductive requires some different definition of these terms or some restriction on their interpretation.

Theorem B also tells us something very reasonable about the function (5):

$$\begin{aligned} s(h, e) \geq 0 \text{ implies } s(h, \bar{e}) \leq 0, \\ s(\bar{h}, e) \leq 0, s(\bar{h}, \bar{e}) \geq 0 \end{aligned} \quad (6)$$

$$\begin{aligned} \text{while } s(h, e) \leq 0 \text{ implies } s(h, \bar{e}) \geq 0, \\ s(\bar{h}, e) \geq 0, s(\bar{h}, \bar{e}) \leq 0 \end{aligned} \quad (7)$$

In words, if e supports h then it counter-supports $\bar{h}$, and also $\bar{e}$ counter-supports h but supports $\bar{h}$. The converse is given by (7).

The above considerations explain, we hope, some of the results obtained by PM, by your correspondents and by us. They do not really go beyond basic facts such as: $h \vee e$ follows deductively from e , $h \vee \bar{e}$ follows deductively from $\bar{e}$, together with probability algebra. In that sense, they are obtained by strict deduction and have nothing to do with induction. It is for these reasons that we have to give the answer "No" to the question in the title.

J. WISE

P. T. LANDSBERG

University of Southampton
Southampton SO9 5NH, UK

1. Popper, K. & Miller, D. *Nature* 302, 687 (1983); 310, 434 (1984).

2. Levi, I. *Nature* 310, 433 (1984).

3. Jeffrey, R. C. *Nature* 310, 433 (1984).

4. Good, I. J. *Nature* 310, 434 (1984).

POPPER AND MILLER REPLY—In 1878, Peirce drew a sharp distinction between "explicative, analytic, or deductive" and "ampliative, synthetic, or (loosely speaking) inductive" reasoning. He characterized the latter as reasoning in which "the facts summed up in the conclusion are not among those stated in the premisses". This was the background to our brief letters (refs 1, 2 in Wise and Landsberg's letter). We identified the ampliative part, relative to any evidence e , of any hypothesis h , with the conditional proposition $h \leftarrow e$; and we proved quite

generally that if $p(h, e) \neq 1 \neq p(e)$, then

$$s(h \leftarrow e, e) = -ct(h, e)ct(e) < 0 \quad (1)$$

where

$$s(x, y) = p(x, y) - p(x)$$

and

$$ct(x, y) = 1 - p(x, y).$$

Accordingly, the ("ampliative") part of the hypothesis h that goes beyond the evidence e is always counter-supported by the evidence. This was our thesis. (The other part, which we wrote $h \vee e$, is of course a deductive consequence of e , and is deductively supported by e .) Both the measures of support and of content are additive:

$$s(h, z) = s(h \vee e, z) + s(h \leftarrow e, z) \quad (2)$$

$$ct(h, z) = ct(h \vee e, z) + ct(h \leftarrow e, z) \quad (3)$$

(Equations (2) and (3) are valid for all h , e and z , and thus for $z = e$.) Wise & Landsberg note that there are other ways of writing $s(h, e)$ in the form $\pm s(x, e) + \pm s(y, e)$; for example

$$s(h, e) = -s(\bar{h} \vee e, e) - s(\bar{h} \leftarrow e, e) \quad (4)$$

This equation (4) is a simplification of their formula (2). It can be obtained by substituting $\bar{h}$ for h in our (2) above, and multiplying by -1 . Obviously, neither $\bar{h} \vee e$ nor $\bar{h} \leftarrow e$ is part of the content of h . Thus we fail to see how our (4)—that is to say, their (2)—can possibly challenge our thesis, or have any bearing whatsoever on it.

KARL POPPER

Fallowfield, Manor Close,
Penn, High Wycombe,
Buckinghamshire HP10 8HZ, UK

DAVID MILLER

Department of Philosophy,
University of Warwick,
Coventry CV4 7AL, UK

Gastric cancer and salivary nitrate and nitrite

SIR — Forman *et al.*¹ reported lower salivary and nitrite levels in healthy subjects from areas of Great Britain with high gastric cancer incidences, compared with subjects from areas of that country with low incidences of this cancer. These unexpected findings were used to argue against the theory^{2,3} that the intra-gastric formation of carcinogenic *N*-nitroso compounds (NOC) from nitrite is involved in the aetiology of gastric cancer. Although this theory remains controversial, I have three objections to the suggestion that the results presented by Forman *et al.* disprove the NOC theory.

(1) A lower nitrate intake was reported¹ in the diet of the high-risk subjects, which accounted for their lower levels of salivary nitrate and nitrite. Since most dietary nitrate (76–89% in this study) arises from vegetables, which supply much of our dietary vitamin C, the high-risk subjects

probably ingested less vitamin C than did those from the low-risk areas. Hence the high-risk subjects could have produced more intra-gastric NOC, despite their lower salivary nitrite levels. The relevant property of each vegetable may be the ratio of vitamin C to nitrate content, and not nitrate content alone⁴. The intake of fresh fruits and vegetables is negatively correlated with gastric cancer incidence², suggesting that the critical components here are vitamin C and other nitrosation inhibitors.

(2) The same argument applies to socio-economic class, since the "lower" classes have relatively high incidences of gastric cancer² and showed lower salivary nitrate and nitrite levels in the reported study¹, but would be expected to consume less vitamin C in fresh fruits and vegetables.

(3) Forman *et al.* reported lower salivary nitrate and nitrite levels in cigarette smokers than in nonsmokers, as did Ladd *et al.*⁵. Both groups attributed this finding to the high level of salivary thiocyanate in smokers. Smokers have raised incidences of various types of cancer, possibly including gastric cancer^{1,2}. Some of these cancers are likely to be induced by NOC in tobacco smoke or produced *in vivo* from smoke constituents⁶. These findings in smokers do not, as suggested¹, discredit the NOC theory because thiocyanate is a powerful catalyst of nitrosation⁷ and we might therefore expect smokers to show an increased NOC formation, despite their lower levels of salivary nitrite.

In 1981 Ohshima and Bartsch⁸ developed a method of measuring the potential for *in vivo* (probably intragastric) nitrosation. They fed nitrate and the amino acid L-proline to volunteers, and detected microgram amounts of the noncarcinogenic nitrosamine, *N*-nitrosoproline, in the 24-hour urines. Using this test, Ladd *et al.*⁵ reported that smokers given nitrate and proline indeed had a greater urinary excretion of nitrosoproline than did non-smokers, despite the smokers' lower salivary nitrate and nitrite levels. Similar nitrosoproline data were reported by Hoffmann and Brunnenmann⁹. These findings prove that salivary nitrite levels cannot be used to predict the potential for intra-gastric NOC formation.

In conclusion, it is interesting that salivary nitrate and nitrite levels follow the opposite trend to that expected from the NOC theory. However, *in vivo* nitrosoproline formation^{5,8,9} offers a more valid test for potential *in vivo* NOC formation and the exposure to nitrosation inhibitors, such as vitamin C, and catalysts, such as thiocyanate, must be taken into account. (The role of vitamin C was addressed in ref. 1, but with a different emphasis.)

SIDNEY S. MIRVISH

*Eppley Institute for Research in Cancer,
University of Nebraska Medical Center,
Omaha, Nebraska 68105, USA*

1. Forman, D., Al-Dabbagh, S. & Doll, R. *Nature* **313**, 620-625 (1985).

2. Mirvish, S.S. *J. natn. Cancer Inst.* **71**, 631-647 (1983).

3. Correa, P., Haenzel, W. & Tannenbaum, S. *Natn. Cancer Inst. Monogr.* **62**, 129-134.
4. *The Health Effects of Nitrate, Nitrite and N-Nitroso Compounds* (Washington, D.C., 1981).
5. Ladd, K.F., Newmark, H.L. & Arslan, M.D. *J. natn. Cancer Inst.* **73**, 83-87 (1984).
6. Hoffman, D., Brunnenmann, R.D., Adams, J.D. & Hecht, S.S. in *N-Nitroso Compounds: Occurrence, Biological Effects and Relevance to Human Cancer* (eds O'Neill, I.K., VonBorstel, Miller, C.T., Long, J. & Bartsch, H.) 743-762 (Oxford University Press, 1984).
7. Boyland, E., Nice, E. & Williams, K. *Food Cosmet. Tox.* **9**, 639-643 (1971).
8. Ohshima, H. & Bartsch, H. *Cancer Res.* **41**, 3685-3662 (1981).
9. Hoffman, D. & Brunnenmann, R.D. *Cancer Res.* **43**, 5570-5574 (1983).

FORMAN *ET AL.* REPLY — In our study, we did not set out to disprove the theory "that the intragastric formulation of carcinogenic *N*-nitroso compounds (NOC) from nitrite is involved in the aetiology of gastric cancer", but rather to test the hypothesis "that an increase in nitrate exposure may represent an increased cancer risk". We agree with Mirvish that the negative relationship we observed between salivary nitrates and nitrites and gastric cancer could be due to an association between the consumption of vitamin C (or other protective factors in vegetables) and the consumption of nitrates and suggested this in our paper. We agree, too, that the increased secretion of thiocyanates by smokers might account for the increased risk of gastric cancer in cigarette smokers that has been reported in several series. We clearly stated in the paper that our results do no more than "weigh against the idea that environmental nitrates and nitrites play a major role in determining the risk of gastric cancer in Britain" and we emphasised that this should not "be taken to imply that nitrate-related *N*-nitroso compound carcinogenesis has no role in the development of gastric tumours". Our conclusion is supported by the recently published results of Beresford showing that in Britain the risk of gastric cancer is negatively correlated with the nitrate content of the water supply in over 200 urban areas (*Int. J. Epidemiol.* **14**, 57-63; 1985) and by our observations on men producing nitrate fertilizers, who are certainly exposed to unusually large amounts of nitrates, and whose mortality from gastric cancer is almost identical with that experienced by other men in the same part of Britain (our work, in preparation).

It is clear that whatever the role of NOC in human carcinogenesis may be, direct exposure to nitrate *per se*, does not appear to be a critical risk factor in the United Kingdom. We agree that *in vivo* nitrosoproline formation has the potential for a useful test of *in vivo* NOC formation and are ourselves conducting more research in this area.

DAVID FORMAN

SAMIM AL-DABBAGH

RICHARD DOLL

*ICRF Cancer Epidemiology
and Clinical Trials Unit,
Gibson Building,
The Radcliffe Infirmary,
Oxford OX2 6HE, UK*

© 1985 Nature Publishing Group

Spreading rate of an Arctic ice shelf

SIR — In connection with problems of creep and *à propos* the work of Doake and Wolff¹ on the flow law for polar ice sheets, Weertman² calls for measurements of the spreading rate of the Ward Hunt Ice Shelf, Ellesmere Island, Arctic Canada. This was an oversight on Weertman's part because, when I was in charge of field operations in the area for the Canadian Defence Research Board, I arranged for such measurements to be made under the supervision of Dr G. Konecny, then of the University of New Brunswick.

The work was performed with a geodimeter during the field seasons 1964, 1965 and 1968, and the data were analysed in detail by Dorner³, who derived a value of $1.17 \times 10^{-4} \text{ yr}^{-1}$ for the spreading rate where the mean thickness of the ice shelf was 38 m. Weertman now agrees that both Doake and Wolff's ice shelf data and Dorner's ice shelf data appear to go against a linear creep law.

G. HATTERSLEY-SMITH

*British Antarctic Survey,
Natural Environment
Research Council,
High Cross, Madingley Road,
Cambridge CB3 0ET, UK*

1. Doake, C.S.M. & Wolff, E.W. *Nature* **314**, 255-57 (1985).

2. Weertman, J. *Nature* **314**, 227 (1985).

3. Dorner, E. *J. Glaciol.* **10**, 211-25 (1971).

Powers of ten correctly expressed

SIR — Claude Liébecq (*Nature* **314**, 586; 1985) thinks that an ordinate with figures ranging from 0 to 10 and labelled "Radioactivity (c.p.m. $\times 10^{-3}$)" is labelled incorrectly if, for example, one of the values included is 5,000 c.p.m. It is indeed labelled incorrectly if one thinks of the label as an equation in the form "Radioactivity = $5,000 \times \text{c.p.m.}$ ". But if, with more logic, one thinks of the words written vertically simply as a label describing what the figures are, then "c.p.m. $\times 10^{-3}$ " is correct.

In other words, instead of substituting "5" for the "c.p.m." in the expression "c.p.m. $\times 10^{-3}$ " (giving the incorrect "Radioactivity = 5×10^{-3} "), one may understand the ordinate to mean "c.p.m. $\times 10^{-3} = 5$ ", or "c.p.m. = 5×10^3 ".

Clearly, the meaning of the ordinate can be read in different ways, and there ought to be agreement on how to read it. I suggest that the words on the ordinate be viewed, as I believe they already are by a majority of my colleagues, simply as a label describing the figures, and not as part of an equation.

PETER LUYKX

*Department of Biology,
University of Miami,
Coral Gables,
Florida 33124, USA*

Prospects for an ageing population

from Jacob A. Brody

The ageing populations of industrialized societies are more likely to be encumbered by increased morbidity than some optimistic arguments would suggest.

OLIVER Wendall Holms, American physician-poet, likened human ageing to the obsolescence of a "one-hoss shay".

Two metaphors of ageing:

Have you heard of the wonderful one-hoss shay,
That was built in such a logical way
It ran a hundred years to a day,
And then, of a sudden, it...

You see, of course, if you're not a dunce,
How it went to pieces all at once,
All at once, and nothing first,
Just as bubbles do when they burst.

End of the wonderful one-hoss shay.
Logic is logic. That's all I say.

In stark contrast are W.B. Yeats' lines in the poem "Sailing to Byzantium":

Consume my heart away; sick with desire
And fastened to a dying animal
It knows not what it is.

HUMAN lifespan, or survival potential¹, cannot be less than 110 to 115 years; some humans have survived that long. It is even possible that lifespan will be found to exceed this estimate, perhaps by a considerable amount. Life expectancy, on the other hand, the average of the years of life from birth or other stated age, is a projection based on accumulated data¹. In the United States, life expectancy is now 71 years for males and 78.3 years for females, an average of about 74 years for both sexes². This discrepancy between life expectancy and lifespan, about 40 years, is a matter of legitimate concern.

What are the prospects for the years ahead? In an influential article published in 1980, J.F. Fries³ postulated that the mean age of death, or the average potential life expectancy, would eventually become 85 years, with 95 per cent of deaths occurring between ages 77 and 93. This consideration is not easily reconciled with current life expectancy in the United States, which at age 85 is 6.3 years and at age 91 remains 4 years⁴. Most other authors⁵⁻⁷ and the Social Security Administration⁸ have therefore concluded that average life expectancy will increase to more than 85 years within a century. Yet, Fries describes his conclusions as: "a set of predictions that contradict the conventional anticipation of an ever older, ever more feeble, and ever more expensive-to-care-for populace. These predictions suggest that the number of very old persons will not increase, that the average period of diminished physical vigour will decrease, that chronic disease will occupy a smaller proportion of the typical lifespan, and that the need for medical care in later life will decrease." This conclusion offers a comfortable solu-

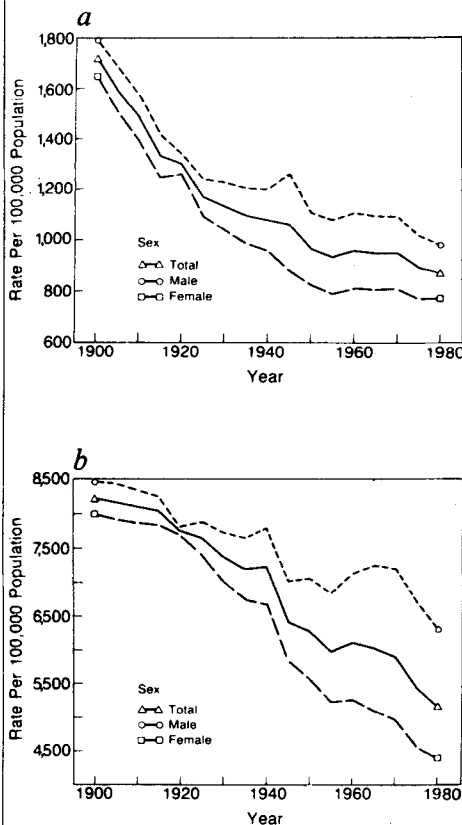


Fig. 1 a, Mortality rate for years 1900-79, by sex (all ages). b, Mortality rates for years 1900-79, by sex (ages 65 and over). Source: National Center for Health Statistics, various volumes of *Vital Statistics of the United States*.

tion as a daunting problem to health policymakers and fiscal planners, but may engender a dangerous optimism. Here I present a less simplified account of the balance between mortality and morbidity in this century and the next.

Mortality reduction

In 1900, only a quarter of the US population survived to the age of 65. Now almost 70 per cent achieve this age, and the proportion of those aged 80 and over is 30 per cent and increasing⁹. Within a very few years, about half the population will live more than 80 years. But the pattern of reduction in mortality has not been smooth over time, or easily explicable. Figure 1a shows that the mortality rate per 100,000 for all ages fell from 1,720 to 868 between 1900 and 1980. Half of the remarkable decrease observed this century had been complete by about 1920. Neither physicians nor social planners were prime movers in the period during which the greatest mor-

tality reduction occurred — improved living environments and better sanitation were the major factors in the decline.

Among those aged 65 and over, mortality had declined by barely 5 per cent by 1920 (Fig. 1b), an indication that the older population was resistant to whatever factors caused the conservation of life in the general population. By about 1945, however, half of the decline in mortality among those 65 and over in this century had been completed. The era between 1920 and 1945 was an historic one in the United States — including the Second World War, the Depression, the beginning of Social Security and Prohibition. While any number of influences may have caused the decline of mortality by 1945, specific medical or social interventions are unlikely to have been major contributors.

During the period from 1945 to about 1968, when it may be thought that medical and social factors should have been reducing mortality rates for the entire population, the overall rate was generally unchanged, with an actual increase in mortality among those aged 65 and over. This seems to have reassured policymakers and planners and, in a sense, lulled them into the false assumption that the maximum potential for human life expectancy had been achieved¹⁰. Yet since about 1968, mortality rates have persistently declined with only minor perturbations (Fig. 1). Declining mortality among those 65 and over has been remarkable, causing a disproportionate accumulation of the oldest and most vulnerable sector of the population.

Can this recent trend be ascribed to medical and social interventions? Undoubtedly they have had some impact, particularly by the application of medical means for the reduction of blood pressure and, in the United States, by the improvement of medical care through Medicare and other health-insurance programmes. There has also been reduction in smoking and some improvements in diet and exercise; while the scientific demonstration of the benefits of diet and exercise is fragile, presumably some good comes from an intelligent and moderate approach to eating and activity. But careful analysis of the data about the powerful determinants of human health and longevity, such as the avoidance of smoking and the use of anti-hypertensive agents, reveals that neither of these influences can explain either the extent or the timing of the decline of mortality in the elderly since 1968¹¹. The major impact of blood-pressure control and

reduced smoking has yet to be fully realized, so there are likely to be even greater gains in life expectancy in future years.

Yet, actuarial arguments such as those used by Fries, as well as the projections of the Social Security Administration, serve as the cornerstone of major policy decisions. My own discomfort with this approach stems from my conviction that we do not have satisfactory answers to the questions of what has brought about this dramatic mortality decline of the last few decades, and of what still sustains it.

In the past 150 years, there have been major shifts in the causes of mortality and morbidity which have profoundly influenced attitudes towards public health and policy. Important diseases have disappeared or diminished, but for the most part we do not know why. Many of the greatest medical triumphs are not easily attributable to our own constructive action¹².

Tuberculosis was the chief cause of death in the United States and northern Europe at the turn of the century, but even by then there had been a very rapid decline in mortality from tuberculosis since 1856 (Fig. 2). The tuberculosis bacillus was first identified in 1882, and specific therapies for tuberculosis were introduced only in the late 1940s. These developments were, therefore, of small importance in the decline of the disease. Similarly, for reasons not understood, deaths from scarlet fever declined precipitously from 1870 to 1900, although the sulpha-drugs and penicillin were not in common use until the Second World War. Moreover, surprising shifts in causes of death are not confined to diseases of infectious aetiology. The rate of mortality from accidents among the elderly has declined rapidly since 1950¹³ as have deaths from stroke and stomach cancer¹⁴, while conversion hysteria, one of Freud's most frequent diagnoses, seems to have vanished.

One phenomenon that has, however, had a great impact on health trends and thus on current attitudes is the remarkable decline since the late 1960s in mortality from ischaemic heart disease¹⁵ which, in my opinion, created a false sense of optimism about the extent to which we are in control of our survival^{11,16}. Heart attacks appear to be a twentieth-century medical chimera. There is no real mention of the classical heart attack or myocardial infarction before 1900: the generally accepted first reference is Herrick's in 1912¹⁷. It is possible but unlikely that the medical profession missed the diagnosis of myocardial infarction or systematically misclassified it; the extraordinarily dramatic nature of a heart attack, with its crushing chest pain and pain radiating down the left arm make it a notable event. Surely our authors and playwrights would have described an event so frequent and dramatic as we now witness, but there are no good descriptions of heart attacks in a literature rich with accounts of gout, epilepsy, apoplexy and

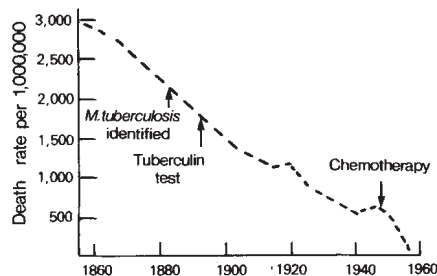


Fig. 2 Respiratory tuberculosis in the total population. Source: Mean death rates from various diseases, England and Wales, 1860-1960. From ref. 45.

tuberculosis. It is probable that acute myocardial infarction was projected into prominence by circumstances arising early in this century, probably associated with the introduction of cigarette smoking and different types and amounts of fat in the diet. Mortality from heart attacks has declined rapidly in the United States since 1968, but this decline has not occurred throughout the world. In Sweden¹⁸ and Japan¹⁹ there has been no comparable decline, although life expectancy is greater than in the United States.

Mortality from heart attacks increased during a period when general mortality was declining. Cause-specific mortality shifted early this century from infectious diseases to chronic diseases and now half of all deaths are caused by heart disease, a large fraction of which result from acute myocardial infarction. After the Second World War, the decline in mortality ended, and was followed by a plateau and then a rise (Fig. 1). This may have resulted from an abatement of the forces that had reduced mortality early in the century, and an increase of mortality from heart attacks. Acute myocardial infarction, however, appears to have been a chance phenomenon which, while causing serious damage, depended for its maintenance and increase on tenuous and complex factors, certainly including smoking and diet. The rapid decline in deaths from acute myocardial infarction began in the late 1960s and it is likely that United States age-specific mortality rates since 1900 would have shown a smooth steady decline over the entire years had it not been for the shifting pattern of mortality from this cause.

So this underlying momentum for life extension seems not to be determined by a decline in cause-specific mortality, but rather to be part of a pattern in which the potential lifespan is ever more closely attained. But concomitantly with increased longevity, there has also been a steady increase in average height of successive generations for more than a century. Gains in growth probably result from the reduction in infectious diseases and from better nutrition in the formative years. There is evidence, in some populations, that growth may now have reached a plateau²⁰.

It is hard not to think that there must be a relationship between these phenomena.

Each generation is more closely approaching the human growth potential and this process is apparently reflected in successively-improved trajectories for survival. Therefore, studying the past history of disease and mortality data can yield only fragile guesses about human survival.

The problem is further compounded because 17 per cent of current deaths in the United States occur in people born before 1900, before the era of complete and accurate vital statistics, so we do not have complete data from even a single cohort upon which to base projections. But it is a fact that older people are living longer⁹. Moreover, no other species has ever had our potential for controlling our environment, so that the limits of life expectancy or maximum attainable longevity are not known. There are no empirical data of great value to instruct us on the limits of these unprecedented happenings.

Disability

Life expectancy is increasing at a remarkable rate among the elderly, but are their lives more satisfactory? Scientific documentation of quality of life and life satisfaction is elusive. Present data indicate a modest gain in healthy years lived and a far greater increase in years of compromised physical, mental and social function^{2,21-24}. The net increase in survival translates into more years of ill health.

Logically, greater longevity should be associated with increased years of good health, but critical studies are insufficient. Limited evidence is accumulating in current studies of successive aged cohorts. But there are no data suggesting an age-specific decline in deafness, blindness, hip fracture, osteoarthritis and Alzheimer's disease and related disorders⁹.

We lack methods to measure the prevalence of good or compromised health. A promising measurement of health in terms other than of disease and death²⁷ is "active life expectancy", the endpoint being the loss of independence in daily life. This study documented a substantial decline in active life expectancy between age 65 and 84. Men were shown to die shortly after losing independence, while women lived on with diminishing functional abilities suggesting that, in the future, as male patterns more closely resemble those of females, the number of person-years lived after active life expectancy will further increase. Currently, limitations of activity due to chronic health conditions, in the age groups 65 to 74 years, 75 to 84 years, and 85 years and older are 41 per cent, 51 per cent, and 60 per cent, respectively²⁸. Annual rates of admission to short-stay hospitals per 1,000 population for these age groups are 299, 451, and 507, and the rates of admission to nursing homes per 1,000, for the same age groups, are 15, 68, and 216. If populations were appropriately studied over time, a pattern of increased active life expectancy might emerge.

Time/trend analysis of health status,

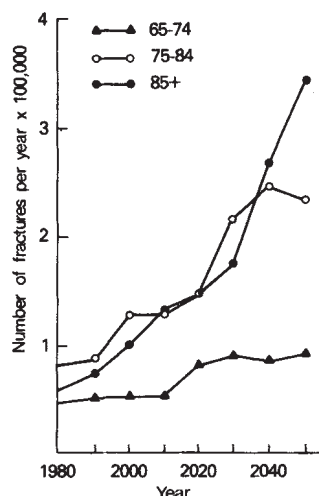


Fig. 3 Projected number of hip fractures annually in the United States by age: 1980-2050. Source: NCHS and US Bureau of Census projections.

however, is unreliable. Colvez and Blanchet²⁹ found a disturbing rise in morbidity and disability since 1957, as did a recent Veterans Administration study³⁰. Wilson³¹, of the National Center for Health Statistics, commenting on the Colvez and Blanchet report, questions the ability "of health statistics to reflect what might really be happening to health status." He offers plausible explanations for the observed increases in morbidity and disability including artificially high results caused by increased sensitivity of the screening instrument, to a change in attitude towards self-reporting as a result of more liberal retirement benefits for health reasons and improved survival of the chronically ill (implying that the increase of morbidity is real). All this merely emphasizes that there are not sufficient data to reach conclusions about whether people are indeed enjoying better health. Increased public health awareness further confounds interpretation of trend analysis. For instance, as education about hypertension increased in the United States, more people had blood pressure determinations which are a key item in a health-status questionnaire, so that the rate of reported hypertension has increased^{29,30}.

Longitudinal studies, in which subjects are regularly examined over time, avoid some of the obvious weaknesses of the sequential and cross-sectional surveys but are few in number and, because of the cost and logistical problems, follow only relatively small cohorts. Yet intelligent health policy rests on such research data. Nobody seriously doubts that the disease burden increases with age. It is essential, however, to monitor morbidity, disability and even the quality of life to determine the trends of health as life expectancy increases.

Population

The dangers of misinterpreting data are matched by those of presuming knowledge that does not exist. The combination of in-

creasing life expectancy and maturation of the "baby-boom" population will produce extraordinary increases in the numbers of older people in the United States. Within about 40 years (by 2025), the population 65 and over will rise from 11.3 to 20 per cent of the total and will comprise about 60 million individuals. By 2050, the population 85 and over will have risen from about 2 million to more than 16 million³. In 10 years, more than half of the people now aged 65 and over will be aged 75 and over. When we now lament that our medical and social systems cannot at present provide for those 65 and over, we forget that it is after age 75 that health problems increase sharply³². A concerned government should focus on the realities of the future, not be seduced by fragile notions that the numbers of very old people will not increase, that impairment of physical vigour will decrease and that the result will be the need for less medical care in later life³.

Three specific examples illustrate physical, mental, and social realities that will occur in the United States unless miraculous advances in science and health management develop in the very near future. In 1980 there were approximately 200,000 people with hip fractures whose median age was 79. Census projections indicate that by the year 2000, before the ageing of the "baby-boom" population, the number will have risen to 330,600 and, by 2050, to more than 650,000, including 340,000 over age 85 (Fig. 3).

In 1980 there were more than 2,000,000 people with Alzheimer's disease and related disorders whose mean age was about 80 years. By the year 2000 the number will rise to 3,800,000 and by 2050 there will be more than 8,500,000, of whom almost 5,000,000 will be age 85 and over (Fig. 4).

There are now about 1.1 million people age 65 and over in nursing homes whose median age is about 80. By the year 2000, the number will rise to 2,200,000 and in 2050 a conservative estimate projects a total of 5,400,000.

Prospects

Unlike the "one-hoss shay", the human body outlives its joints, its cognitive and memory capacities, its vision, its hearing and the many other substrates of disability. While concern is limitless, there is at present no accepted scientific prospect that the net increase in healthy years will exceed the increase of years of dysfunction. We are, however, far short of reaching the potential of better health and quality of life in later years. Two vehicles offer promise: health promotion through the improvement of personal health practices throughout life, and major research efforts to understand and postpone the ageing processes that inexorably move the elderly into an increasingly morbid and disabled state.

The value of health promotion, involving people in self-care and in achieving wholesome lifestyles is obvious, but a crisis of credibility is emerging in the pursuit of

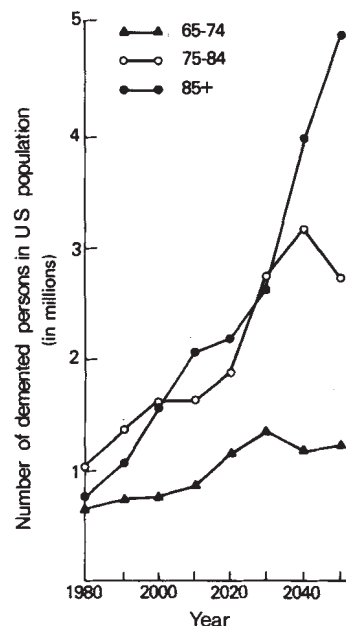


Fig. 4 Projected number of demented persons in the United States by age: 1980-2050. Source: NTA prevalence estimates and US Bureau of Census projections.

this goal. One hazard is overstating knowledge when scientific evidence is weak or lacking. In the face of the growing numbers of old and very old people, we must admit that we do not know causes of, and cannot prevent, most chronic diseases³³; only a fraction of the cancers, heart diseases, strokes, neurological diseases and arthritides are understood, and we know virtually nothing about the causes and prevention of the senile dementias that afflict 20 per cent of people 80 years of age and older³⁴.

Nobody doubts that smoking and excessive drinking are harmful. Data substantiating the importance of effective treatment of high blood pressure are robust, and at least one study has shown that lowering cholesterol levels is beneficial³⁵. Data relating to exercise, weight control, special diets and salt restriction in the absence of hypertension are much less persuasive. Nor do we know how great would be the reduction in morbidity and disability if we produced a population of paragons. Indeed, with almost 70 per cent of the US population now surviving their 65th year, data of Branch and Jette³⁶ indicate almost no association between most of the commonly advocated personal health practices and mortality among the elderly. To maintain our credibility, we have an obligation to admit our ignorance of causation and our propensity to engender guilt and anxiety among the elderly in areas in which the benefits of specific preventive techniques are not scientifically documented¹⁶.

Another area of health promotion that requires rethinking is behaviour modification through health education. Although considerable benefits may be achieved by altering lifestyles, the results so far have been uneven and equivocal^{37,38}. There may

be a systematic obstacle in health education related to an individual's educational level. The classic US study by Kitagawa and Hauser shows a distinct survival advantage among the better educated³⁹ and is supported in the current literature^{40,41}. A recent British investigation reported similar findings⁴², noteworthy because medical services have been available free in Britain since 1948.

Education (and its concomitants, all of which are harder to measure) has a powerful effect on life expectancy. People with minimal education may not be reached by the available means of health education. Elimination of smoking, moderation of alcohol consumption, diet control and other attempts at modification of lifestyle will have by-passed this less well educated group, which continues to inflict needless self-harm. We need a new direction in behaviour modification through health education, with the primary target being those with fewer years of education. We may be wasting much of our resources by failing to modify clearly harmful health practices in this inaccessible high-risk subpopulation.

In research, it is important that we should place emphasis on conditions that become more prevalent with ageing, such as blindness, deafness, Alzheimer's disease, acute and chronic heart disease, diabetes, and hip fracture, and that we should develop mechanisms to postpone them. This approach has increasing value as life expectancy reaches or passes 85 years, as illustrated with hip fractures and the associated osteoporotic process. Our data⁴³ indicate that hip fractures start to rise in frequency in the early forties and thereafter increase exponentially, doubling every 6 years (Fig. 5). If research on bone metabolism could provide a mechanism for delaying osteoporosis so as uniformly to postpone the onset of the exponential rise in hip fracture, the incidence of this condition would be reduced by 50 per cent⁴⁴. Research directed at the precursors of disability is the most effective tool to prevent chronic disease and attendant disabilities.

The most urgent need is of an understanding of human natural history beyond age 65, using the emerging data on elderly populations. The increase in life expectancy among the elderly is unprecedented in human history and probably relates to longer survival trajectories for successive cohorts rather than to specific disease control. The causes of the systematic alteration between 1900, when 25 per cent of the population survived their 65th year, and 1980, when almost 70 per cent did so, demands explanation.

Meanwhile, in our state of ignorance, making assumptions with major policy implications is dangerous. When we speculate in good faith, but on a basis of soft data and with projections that human history has already denied, we run the risk of ignoring Hippocrates' basic admonition that

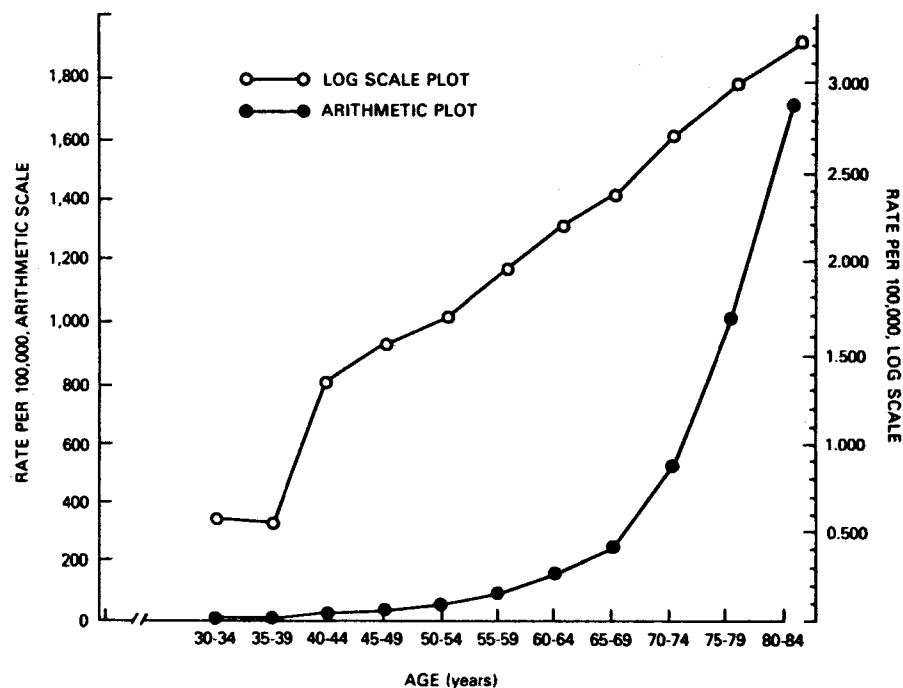


Fig. 5 Age-specific hip fracture incidence among white women. Source: National Hospital Discharge Survey 1975 to 1979.

"the physician must be able to tell the antecedents, know the present and foretell the future — must meditate these things, and have two special objects in view with regard to diseases, namely, to do good or to do no harm."

Parts of this paper were presented at a University of Connecticut Symposium and will appear in the *Philosophy and Medicine Series* by D. Reidel Publishing Company. □

Jacob A. Brody is Associate Director for Epidemiology, Demography, and Biometry Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20205, USA.

- Schneider, E.L. & Brody, J.A. *New Engl. J. Med.* **309**, 854-856 (1983).
- Taubert, C.M. *Current Population Reports Series P-23*, No. 128, (US Government Printing Office, Washington, D.C., 1983).
- Fries, J.F. *N. Engl. J. Med.* **303**, 130-135 (1980).
- Spencer, G. *Curr. Popul. Rep. Series P-25*, No. 952, (US Government Printing Office, Washington, D.C., 1984).
- Manton, K.G. *Milbank Memorial Fund Quarterly/Health and Society* **60**, 183-244 (1982).
- Brody, J.A. *J. Am. Geriatr. Soc.* **30**, 681-683 (1982).
- Rosenwajke, I., Yaffe, N. & Sagi, P.C. *Am. J. publ. Hlth* **70**, 1074-1080 (1980).
- Faber, J.F. & Wilkin, J.C. *Social Security Administration, Office of the Actuary, Actuarial Study No. 85*, SSA Pub. No. 11-11532, (US Department of Health and Human Services, 1981).
- Brody, J.A. *Publ. Hlth Rep.* **99**, 468-475 (1984).
- The Change in Mortality Trends in the U.S.* Series 3, No. 1, (US Government Printing Office, Washington, D.C., 1964).
- Stallones, R.A. *Scient. Am.* **243**, 53-59 (1982).
- Ewbank, D. & Wray, J.D. in *Public Health and Preventative Medicine* 11th Edn (ed. Maxcy-Rosenau) 1504-1519 (Appleton-Century-Crofts, New York, 1980).
- Brody, J.A. & Brock, D.B. in *Handbook of the Biology of Aging* 2nd Edn (eds Finch, C.E. & Schneider, E.L.) (Van Nostrand, New York, 1984).
- Brock, D.B. & Brody, J.A. in *Principals of Geriatric Medicine* (eds Andres R., Bierman, E.L. & Hazzard, W.R.) 51-71 (McGraw-Hill, New York, 1984).
- Proceedings of the Conference on the Decline in Coronary Heart Disease Mortality* (eds Havlik, R.J. & Feinleib, M.) NIH Publ No. 79-1610, (US Government Printing Office, Washington, D.C., 1979).

- Thomas, L. in *The Medusa and the Snail*, 165-166, (Viking, New York, 1979).
- Herrick, J.B. *J. Am. med. Ass.* **59**, 2015-2020 (1912).
- Welin, L. *et al. Lancet* **i**, 1087-1089 (1983).
- Kimura, N. *Prev. Med.* **12**, 222-227 (1983).
- Abraham, S., Johnson, C.L. & Najjar, M.F. *Vital and Health Statistics* No. 14, (US Government Printing Office, Washington, D.C., 1977).
- Feinleib, M. in *Proceedings of the Second Conference on the Epidemiology of Aging* (eds Haynes, S.G. & Feinleib, M.) 359, NIH Pub. No. 80-969, (US Government Printing Office, Washington, D.C., 1980).
- Siegel, J.S. in *Proceedings of the Second Conference on the Epidemiology of Aging* (eds Haynes, S.G. & Feinleib, M.) 313, NIH Pub. No. 80-969, (US Government Printing Office, Washington, D.C., 1980).
- Gruenberg, E.M. *Milbank Memorial Fund Quarterly/Health and Society* **55**, 3-24 (1977).
- Kramer, M. *Acta psychiat. scand.* **62**, 382-396 (1980).
- Svanborg, A. *Nordisk Gerontologisk Tidsskrift, Norsam-Nyt*, Suppl. sid. 21-22, 1982.
- Branch, L.G. *Center for Survey Research Monograph*, (Boston University of Massachusetts-Boston, 1977).
- Benfante, R., Reed, D. & Brody, J.A. *J. chronic Dis.* (in the press).
- A Chartbook of the Federal Council on Aging*, DHHS Publ No. (OHDH) 81-20704 (1981).
- Colvez, A. & Blanchet, M. *Am. J. publ. Hlth* **71**, 464-471 (1981).
- Survey of Aging Veterans, A Study of the Means, Resources and Expectations of Veterans Aged 55 and Over* RSM 70-84-3, (US Government Printing Office, Washington, D.C., 1984).
- Wilson, R.W. *Am. J. publ. Hlth* **71**, 461-463 (1981).
- Brody, J.A. & Foley, D.J. in *The Teaching Nursing Home: A New Approach to Geriatric Research, Education, and Clinical Care* (eds List, L., Ory, M., Schneider, E.L., Wendland, C. & Zimmer, A.) (Raven, New York, 1980).
- Brody, J.A. *J. Am. Geriatr. Soc.* **30**, 681-683 (1982).
- Brody, J.A. *Natn. Forum* **62**, 4-5 (1982).
- Hjerman, I., Velve Byre, K., Holme, I. & Leren, P. *Lancet* **ii**, 1303-1310 (1981).
- Branch, L.G. & Jette, A.M. *Am. J. publ. Hlth* **74**, 1126-1129 (1984).
- Kirscht, J.P. *Am. J. publ. Hlth* **73**, 134-135 (1983).
- Mazzuca, S.A. *J. chron. Dis.* **35**, 521-529 (1982).
- Kitagawa, E. & Hauser, P.M. *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*, (Harvard University Press, 1973).
- Data from the National Health Survey*, Ser. 15, No. 1 (National Center for Health Statistics, Hyattsville, Maryland, 1981).
- Ruberman, W., Weinblatt, E., Goldberg, J.D. & Chaudhary, B.S. *J. chron. Dis.* **36**, 151-160 (1983).
- Hollingsworth, J.R. *J. Hlth Soc. Behav.* **22**, 268-263 (1981).
- Farmer, M.E., White, L.R., Brody, J.A. & Bailey, K.R. *Am. J. publ. Hlth* **74**, 1374-1380 (1984).
- Brody, J.A., Farmer, M.E. & White, L.R. *Am. J. publ. Hlth* **74**, 1397-1398 (1984).
- Galen in *Great Books of the World* (ed. Hutchins, R.M.) (46 Encyclopaedia Britannica, Inc., Chicago, 1952).

Mass distribution in the galactic centre

M. K. Crawford, R. Genzel, A. I. Harris, D. T. Jaffe, J. H. Lacy*, J. B. Lugten, E. Serabyn & C. H. Townes

Department of Physics, University of California, Berkeley, California 94720, USA

New infrared and submillimetre spectroscopic measurements of the gas dynamics in the central 10 pc of our Galaxy make a convincing case that the mass distribution at the centre of the Galaxy is more concentrated than a spherical isothermal stellar cluster. The measurements fit a point mass of $\sim 4 \times 10^6 M_{\odot}$ but are also consistent with a cluster where stellar density decreases with radius (R) at least as fast as $R^{-2.7}$, or a combination of a point mass and a stellar cluster. The dynamical information combined with previous 2- μm observations favour a large point mass, which is presumably a massive black hole.

THE possible existence of massive black holes at the nuclei of galaxies is an important long-standing question. The lack of spatial resolution has limited the work in external galaxies, and the observational difficulty in establishing a reliable answer to this question even in our own Galaxy is quite formidable. The centre of our Galaxy is obscured visibly, and material along the line-of-sight to the centre hampers radio investigations. Here we discuss new information on the gas dynamics in the galactic centre, available from spectroscopy in the infrared and submillimetre regions. This new information allows a determination of the rotational velocity of the interstellar material between ~ 0.5 and 10 pc from the centre and gives some discrimination between possible mass distributions. The observations entering into the analysis comprise a large body of data which adds considerable detail to our knowledge of the velocity and mass distribution in the central 10 pc of our Galaxy. These data will be presented and discussed in detail elsewhere¹⁻⁴. Here, we provide a brief summary of our present understanding of the galactic centre based on these new results. Distance to the galactic centre is assumed to be 10 kpc.

Previous observations of inner 10 pc

Figure 1 indicates the distribution of interstellar gas and dust in the inner 10 pc of our Galaxy^{2,5}. High-resolution Very Large Array (VLA) maps⁵⁻⁷ show that much of the free-free radio continuum emission from Sgr A West, the radio source in the galactic centre, is distributed in large-scale filamentary structures which may be spiralling into or out of the galactic centre, or orbiting around it⁵⁻⁷. There is also an unusual radio point source (Fig. 1; refs 8, 9), whose position almost exactly coincides with the infrared source IRS16 which is very close to if not at the dynamical centre of the Galaxy^{10,11}. Spectroscopic measurements of the ionized gas show hydrogen and helium with a velocity range of $1,500 \text{ km s}^{-1}$ within 3 arc s of IRS16 (refs 12, 13). At 10–20 arc s from IRS16, measurements of singly ionized neon¹⁴ and of hydrogen recombination lines¹⁵ indicate substantially lower velocities. A statistical analysis based on the velocity dispersion of Ne II-emitting clouds as a function of distance from IRS16 implies that a point mass of $\sim 3 \times 10^6 M_{\odot}$ combined with an isothermal stellar cluster (density $\propto R^{-2}$, where R is the radius) is a more probable mass distribution for the centre than a R^{-2} stellar cluster, but the level of significance of this result is not high¹⁴.

Far-infrared continuum observations indicate that a ring or disk of neutral dust and gas surrounds the centre from a radius of ~ 2 pc outward and that inside this region dust and gas is of substantially lower average density¹⁶. These data (Fig. 1) imply

that the neutral dust disk is oriented approximately in the plane of the Galaxy such that two lobes representing the inner edge of this disk appear on either side of the ionized region, apparently encompassing it. Far-infrared measurements of the O I fine structure line at $63 \mu\text{m}$ show that this disk rotates around the galactic centre, with an axis approximately that of general galactic rotation¹⁷. Emission from H_2 molecules at $2 \mu\text{m}$ indicates that the disk contains hot ($\sim 2,000 \text{ K}$) gas, which is probably shock-excited¹⁸. Radio observations of the 21-cm H line and the 1-0 rotational line of CO show that cooler material in the disk extends to ≥ 6 pc from the centre¹⁹.

New data on galactic centre

New spectroscopic results provide a much more detailed dynamic picture of the ionized gas features by a more complete examination of Doppler shifts of the $12.8\text{-}\mu\text{m}$ Ne II fine-structure line. A much clearer picture of the velocity distribution in the neutral disk out to ~ 10 pc from the galactic centre comes through mapping of various far-infrared lines, including: (1) fine-structure lines of O° in the ground state (at 63 and $146 \mu\text{m}$); (2) the fine-structure line of C^{+} at $158 \mu\text{m}$; and (3) the $J=7 \rightarrow 6$ rotational transition of CO at $372 \mu\text{m}$. The O I and C II data were taken on the Kuiper Airborne Observatory with a Fabry-Perot spectrometer^{2,3}, the Ne II line with a Fabry-Perot-grating spectrometer using the IRTF telescope on Mauna Kea, Hawaii¹, and the $7 \rightarrow 6$ transition of CO with a newly developed heterodyne spectrometer, also at the IRTF⁴.

Ionized streamers. New spectra of the Ne II line were taken at 110 different locations with $1.5\text{--}6\text{-arc s}$ resolution, principally along the structures in the VLA radio continuum map. The questions of interest were whether the velocities of the filamentary structures are continuous and characteristic of material in systematic motion, and which arms of the various streamers are parts of the same dynamic structures. The results and their interpretation are discussed in detail in ref. 1 (see also ref. 15) and are summarized here.

There are at least two long, coherent streamers of ionized gas. Their velocities and spatial patterns suggest that the gas follows simple orbits around the galactic centre, although there are some deviations from a smooth velocity pattern. As proposed by Lo and Claussen⁷, the arc of ionized gas extending from ~ 30 arc s south of IRS16 to ~ 15 arc s west of the centre continues further north to ~ 30 arc s north of the centre (the 'western arc' in Fig. 1). Velocities of the 16 spectral-line centres at approximately equally spaced positions along this $\sim 180^{\circ}$ arc vary linearly with declination from -100 to $+100 \text{ km s}^{-1}$, with more-or-less random deviations of $\sim 7 \text{ km s}^{-1}$ r.m.s. This corresponds to what one expects from the projection of a rotating circle of radius 1.7 pc and rotational velocity of 110 km s^{-1} made up, however, of individual clouds that are not following precisely the same

* Present address: Astronomy Department, University of Texas, Austin, Texas 78712, USA.

orbit. This rotational velocity allows for the observed ellipticity of the arc, which makes the axis of this circle tilted with respect to our line of sight by $\sim 70^\circ$. Its extremes lie at the inner edge of the neutral dust and gas disk detected in far-infrared continuum and line radiation^{16,17}. As the rotational velocity and projected linear velocity gradient agree with those of the neutral gas at the inner edge of the neutral disk, the western arc is probably the ionized inner surface of the disk. The rather clear-cut circular orbit of the western arc allows a determination of the enclosed mass. If the mass distribution within 1.7 pc is spherical, it contains $4.7 \pm 1 \times 10^6 M_\odot$. That the circular orbit is ionized and hence visible in Ne II over only about half of its circumference may result from shadowing of a central ionizing source. The shadowing cloud may be, for example, the 'northern arm', which could interrupt ultraviolet radiation west of the centre.

The northern arm (Fig. 1) is another long structure which appears to be a continuous kinematic feature, with velocities varying from 137 km s^{-1} on the northern end to about zero at the declination of the centre. The average change of velocity with distance along this feature is three to four times greater than along the western arc. Hence, this gas feature is closer to the dynamical centre. The velocities in the northern arm reach a maximum of 137 km s^{-1} at a projected distance of 15 arc s (0.75 pc) from IRS16, suggesting that $\sim 4 \times 10^6 M_\odot$ would have to be enclosed within this radius if this feature were a circular orbit. However, the significant curvature of the velocity curve of the arm indicates that the gas feature is not at constant radius, so that a simple circular orbit analysis is not appropriate. A quantitative examination of the velocities and positions shows that the ionized gas in the northern arm is well described by an eccentric orbit about a point mass, such as a parabola, hyperbola or elongated ellipse. Deviations from such a fit are $\sim 2 \text{ km s}^{-1}$ r.m.s., and velocities seem to vary continuously, suggesting an elongation of a single cloud. The inferred orbit passes within 0.5 pc of the central mass of $3.5 \times 10^6 M_\odot$ located within a few arc s of IRS16. The other model which is of particular interest here is an orbit in an isothermal cluster; such an orbit fitting the measurements of the northern arm still requires a mass of $\sim 3.4 \times 10^6 M_\odot$ within a radius of 0.5 pc; hence a mass $\sim 11.5 \times 10^6 M_\odot$ would lie within 1.7 pc, substantially exceeding the mass observed within that radius. It follows that most of the mass contained inside the outer orbit at $R = 1.7 \text{ pc}$ (the western arc) is also within $\sim 0.5 \text{ pc}$ of the centre. This mass distribution can result from either a central point mass or a stellar cluster more concentrated than an isothermal cluster, with density falling off steeper than $R^{-2.7}$.

Neutral material. We have examined the velocities of the neutral material extending from $\sim 30 \text{ arc s}$ to several minutes of arc, or $\sim 10 \text{ pc}$ from the galactic centre in the far-infrared and submillimetre regions with 30–60-arc s angular resolution. The most recent and improved O I measurements are described in detail elsewhere by R.G. *et al.*², the C II results by J.B.L. *et al.*³ and the CO $J = 7 \rightarrow 6$ results by A.H. *et al.*⁴.

The relative intensities of the two O I fine-structure lines and the C II line, as well as the relative intensities of various CO rotational lines^{2,4}, show that most of the neutral material is in clumps with a hydrogen density near 10^5 cm^{-3} and a gas temperature near 300 K. The total mass of the gas in the neutral disk within $\sim 5 \text{ pc}$ is a few times $10^4 M_\odot$, and $< 10\%$ of the volume is filled by high-density gas. Maps of intensities of these lines show that the galactic centre gas disk is tilted by $\sim 20^\circ$ with respect to the galactic plane and may be quite thin (thickness $\leq 1 \text{ pc}$).

The velocities of all far-infrared and submillimetre emission lines are blueshifted at negative galactic longitudes and redshifted at positive longitudes, with a sharp transition at the centre position. The distributions of integrated line intensities and absolute values of the velocities as a function of longitude offset are very similar on either side of the centre. Rotation is the simplest gas motion fitting this symmetrical distribution and the other observed characteristics of the data. From the orienta-

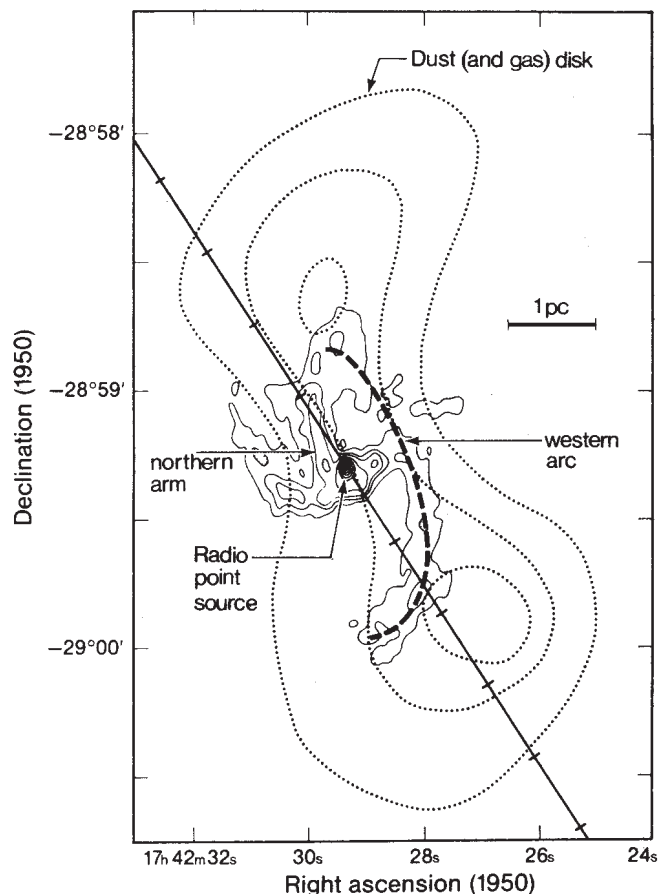


Fig. 1 Sketch of gas in the central 10 pc of the galactic centre. A VLA map of the 2-cm thermal radio continuum emission at a resolution of 2.5 arc s^2 is indicated by the solid contours. Dotted contours give the far-infrared continuum emission from warm dust^{2,16}. The galactic plane is marked by diagonal line with ticks every pc. In addition to the filaments marked, the bar is the structure stretching from north-west to south-east $\sim 10 \text{ arc s}$ on either side of the radio point source.

tion of the disk of gas, noted above, and from the velocities measured, the rotation axis must be near that of the general galactic rotation. The velocity at $R = 1.7 \text{ pc}$ determined from the O I and CO measurements is in excellent agreement with the orbital velocity of the western arc. The spectra indicate that velocities decrease with increasing distance from the centre. The velocities at the peaks of the profiles are $\pm 70 \text{ km s}^{-1}$ at $R \approx 2 \text{ pc}$ on respective sides of the centre, and decrease over the next several parsecs to $\pm 50 \text{ km s}^{-1}$ at 6 pc and $\pm 27 \text{ km s}^{-1}$ at 10 pc. Beyond $\sim 10 \text{ pc}$, the velocity patterns become less clear because the C II lines associated with the dense and warm material close to the centre become weaker and may be contaminated by radiation from other sources along the line of sight. A spectrum of the 1-O CO line at $R = 4.5 \text{ pc}$ (ref. 19) and two Ne II profiles taken at $R = 2.7$ and 5 pc (ref. 20) also support the fall-off of rotational velocity between 2 and 10 pc. The rotation curve is consistent with a 'keplerian ($R^{-1/2}$)' fall-off resulting from orbits around a point mass, and excludes an isothermal cluster.

In addition to rotation, the neutral disk exhibits some non-systematic motion. For example, lines in many locations are as broad as 100 km s^{-1} , and there is a separate blueshifted cloud ($v = -20$ to -80 km s^{-1} with respect to the local standard of rest) away from the galactic plane north-west of the centre. Nevertheless, for distances greater than 1.7 pc, the measurements give clear evidence of an average rotation. At least some of the irregular motions result from a turbulent velocity dispersion of the gas clumps. In the most central region, O I and CO lines show still larger widths, corresponding to velocities detec-

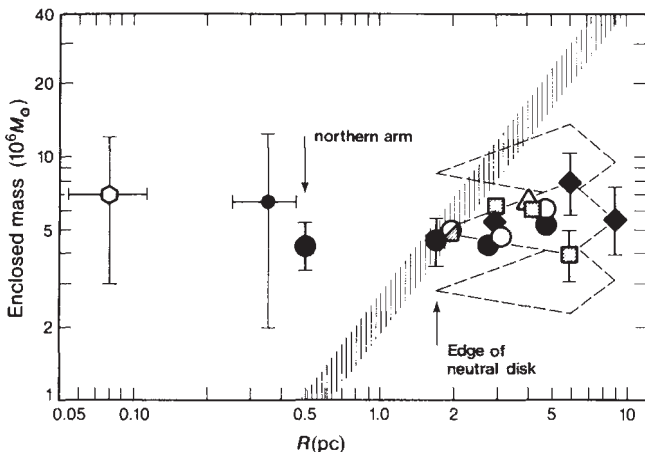


Fig. 2 Enclosed mass as a function of radius from IRS16. The mass distribution is derived from several measurements of the kinematics of neutral and ionized gas discussed in the text. With the exception of the mass enclosed within the northern arm from Serabyn and Lacy's fitting of eccentric orbits, all masses were derived assuming circular orbits: $M_{\odot} = v^2 R / G$. Typical error bars are indicated. Velocities in the neutral disk were corrected for an angle between the rotation axis and line of sight (inclination) of 69° . Rotational velocities of the neutral gas in the disk were derived from the terminal velocities (velocity at half maximum), corrected for a local 'turbulent' velocity dispersion of FWHM 50 km s^{-1} . The large dashed diamonds above and below the measured points show where these data on the neutral gas would be without such a correction for non-circular motions, and where they would be if line centroids rather than edges were taken. Relative positions of the points within the dashed diamonds would be essentially unchanged. He I, $2.1 \mu\text{m}$ ($\circ$); Ne II, $12.8 \mu\text{m}$ ($\bullet$); O I, $63 \mu\text{m}$ ($\circ$); C II, $158 \mu\text{m}$ ($\blacklozenge$); CO 7 $\rightarrow$ 6, $370 \mu\text{m}$ ($\boxtimes$); CO 1 $\rightarrow$ 0, 0.26 cm ($\triangle$). The Ne II data are from refs 1, 14, 20; O I data from ref. 2; C II from ref. 3; CO 7 $\rightarrow$ 6 from ref. 4; and CO 1 $\rightarrow$ 0 from ref. 19. The $2\text{-}\mu\text{m}$ He I data (from ref. 13) indicate that velocities of $\pm 750 \text{ km s}^{-1}$ occur in a region within a radius of 1–2 arcs, leading to the cross at $R \sim 0.08 \text{ pc}$. The shaded diagonal bar indicates the mass distribution of a cluster with $4.7 \times 10^6 M_{\odot}$ within a radius of 1.7 pc and a radial dependence of enclosed mass $R^{1.3}$ (from the $2\text{-}\mu\text{m}$ light distribution; refs 23, 24).

ted earlier for Ne II and indicating that there is neutral material in many of the ionized clumps. Normally, the best estimate of the rotational velocities in an edge-on disk comes from the extreme ('terminal') velocities of the profiles. In the galactic centre, these edges are significantly broadened by the turbulent motions (Δv_{turb} (FWHM) $\sim 30\text{--}50 \text{ km s}^{-1}$). We correct for this broadening in our estimates of the rotational velocities. As indicated in Fig. 2, if turbulence of this magnitude is instead added to the average rotational velocity to calculate average kinetic energy, the mass contained inside a given radius would be increased for the neutral gas measurements, but the fractional rate of change of mass with distance in this region would not be affected. This general result for the fractional rate of change of mass results from the observed decrease in velocity with distance, which makes it relatively insensitive to details of line widths and turbulence.

Discussion

Figure 2, which shows the mass inside of various radii, gives an overview of the mass distribution in the galactic centre. We have derived the masses from the measured velocities assuming a spherical mass distribution. The maximum influence on the derived rotational velocities of corrections for the noncircular motions in the neutral disk are schematically indicated by large dashed diamonds. We choose the points as placed on Fig. 2, using the corrections stated in the legend, as giving the most likely mass values. For other choices, points within the central dashed diamond would approximately maintain their relative

positions but be moved with this diamond to its upper or lower positions, depending on one's choice of corrections. In addition to the estimates for the disk, 'western arc' and 'northern arm', we have also indicated a mass estimate from the velocity dispersion of eight Ne II clouds within 8 arcs of IRS16 (from the earlier work by Lacy *et al.*¹⁴). The line-of-sight velocity dispersion of the inner Ne II clouds, most of which are located in the 'bar' region, is $\sigma_v = 145 \pm 50 \text{ km s}^{-1}$ (ref. 14). We can estimate a mass from the data in Fig. 2 using the virial theorem (enclosed mass $= 3\sigma_v^2 R / G$; see ref. 14), which assumes the gas velocities represent a steady-state distribution. (G is the gravitational constant.) This estimate is about the same as the mass determined from the highest velocity of these clouds ($\pm 260 \text{ km s}^{-1}$) assuming circular motion. Finally, a mass estimate within 1–2 arcs of IRS16 originates from the total width of He I and H I lines measured by Geballe *et al.*¹³, with the assumption of circular motion. In contrast to the other measurements discussed above, it is not certain that the innermost Ne II and He I/H I velocities represent circulation or any simple orbits. The Ne II clouds in the bar region may be falling into the centre⁷, and the He I/H I motions may result from an expanding wind from IRS16 (ref. 13). Hence, these data should be given less weight in the derivation of the mass distribution.

Taken together, the large body of data shown in Fig. 2, measured with different methods and subject to different uncertainties and systematic effects, make a convincing case that at small radii the apparent mass approaches a constant.

The measurements are consistent with a point mass of $4 \pm 1 \times 10^6 M_{\odot}$, or a combination of a $3\text{--}4 \times 10^6 M_{\odot}$ point mass and a small stellar cluster ($1\text{--}2 \times 10^6 M_{\odot}$ within 2 pc and $5\text{--}10 \times 10^6 M_{\odot}$ within 10 pc). They are also consistent with a stellar cluster and no point source if the stellar density decreases as fast or faster than $R^{-2.7}$, and thus exclude an isothermal (R^{-2}) distribution. Some globular clusters and elliptical galaxies have radial density profiles as steep as $R^{-2.7}$ or steeper²¹. Concentrated globular clusters such as M15, for example, have radial density profiles outside their core radii ($\sim 0.01\text{--}1 \text{ pc}$) which approximately follow a power law $R^{-\alpha}$ with $2.4 \leq \alpha \leq 2.8$ (ref. 22). Such a gradient is consistent with the present data and the main difference between the present case and globular clusters would then be the higher central density and smaller core radius ($\sim 0.01\text{--}0.05 \text{ pc}$) of the galactic centre mass distribution.

However, additional information on the stellar distribution in the galactic centre, available from observations of the $2.2\text{-}\mu\text{m}$ stellar surface brightness distribution^{23,24}, makes it unlikely that such a concentrated stellar cluster dominates the mass in the central 2 pc of the Galaxy. The distribution of $2.2\text{-}\mu\text{m}$ surface brightness drops off with $R^{-0.7}$ between 0.05 and 1.5 pc radius from IRS16 (ref. 24), and with $R^{-0.8}$ between 1 and 30 pc (ref. 23) implying a density approximately proportional to $R^{-1.75}$. There is no sign of a flattening in the inward increase of surface brightness, even at the smallest scales observable ($\leq 0.05 \text{ pc}$; ref. 24). The slope of the $2.2\text{-}\mu\text{m}$ emission line is less steep than that of a cluster in isothermal equilibrium, and certainly not steep enough to explain the spectroscopic velocity measurements. The shape of the $2.2\text{-}\mu\text{m}$ light distribution combined with the dynamical information thus suggests strongly that the stars distributed similar to the $2.2\text{-}\mu\text{m}$ light do not contain the major fraction of the mass within 3 pc. Therefore, if the $2.2\text{-}\mu\text{m}$ light distribution in the centre is a measure of the total mass distribution of stars, the mass concentration at the galactic centre probably results from a central point mass of $\sim 4 \times 10^6 M_{\odot}$, most likely in the form of a single massive black hole. Note that the particular radial distribution observed for the stellar cluster would be a natural consequence of a black hole that is more massive than the stellar cluster out to a given radius. Theoretical calculations show that a steady-state balancing between infall of stars towards such a hole and tidal disruption by the hole gives a stellar density distribution proportional to $R^{-1.75}$, similar to what is observed in the galactic centre²⁵.

There are several reasons why the evidence for a highly concentrated mass at the galactic centre surrounded by a less

massive stellar cluster could be misleading. For example, the 2- μm light distribution may not reflect the mass distribution of the galactic centre stellar cluster. The stellar mass distribution may be more concentrated towards the centre than indicated by the 2- μm light if the stars emitting at 2 μm do not represent a major fraction of the stellar mass. However, the 2- μm light distribution is believed to measure the distribution of K and M giants and supergiants; these would normally be the more massive members of the cluster and should be more concentrated than the cluster as a whole.

A spheroidal or triaxial cluster would have velocities that decrease with distance from the centre rather than the simpler behaviour associated with spherical clusters. In the central plane, however, where velocity measurements are considered in the present study, although the ionized clouds are asymmetrically arranged, they contain little of the total mass and the distribution of gas around the galactic centre is rather symmetrical in both density and velocity. Hence, any extreme ellipticity in the stellar cluster that would give the observed velocity distribution seems improbable. A mild ellipticity of the mass distribution would not change the results qualitatively, but would only decrease the total amount of calculated mass. In the case of the northern arm, one might wonder whether some broadly distributed gas, providing friction to the orbit of any streamer of gas, can produce the curvature and accelerations observed. Quinn and Sussman²⁶ have proposed such a distribution of hot gas as an explanation for the morphology of the ionized streamers in the galactic centre. The Ne II observations of the western arc and the northern arm, however, exclude the specific model proposed¹, because the measured velocities are substantially different from those predicted by the model. Any more general friction-induced infall model is only possible if it can explain both the almost undisturbed circular orbit of the western arc and the presence of clouds moving as fast as 260 km s⁻¹ (ref. 14) very near the galactic centre.

If a black hole accounts for $4 \times 10^6 M_\odot$, the measurements in Fig. 2 allow a stellar cluster mass of only $1-2 \times 10^6 M_\odot$ within $R = 2$ pc. The total luminosity of the cluster within this radius, which is deduced by assuming the 2- μm radiation represents a distribution of 4,000 K giant stars, is $\sim 4 \times 10^6 L_\odot$ (refs 23, 24). Hence, the mass to luminosity ratio of the cluster is $\leq 0.5 M_\odot/L_\odot$, which is much lower than typical for old clusters ($M_\odot/L_\odot \sim 1.5-5$; ref. 21). It may indicate relatively recent star formation activity or a lack of low mass stars. The total luminosity of the region inside $R = 1$ pc deduced from the ionizing flux or the longer-wavelength infrared continuum radiation is $\sim 2 \times 10^7 L_\odot$ (ref. 16), and hence if all the mass were assumed to be stars, the ratio $M_\odot/L_\odot$ would be very low.

Related observations

The turbulent and non-circular motions of the ionized and neutral gas in the galactic centre are significantly larger than in other dense clouds in the Galaxy or than one would expect in equilibrium. It is likely that this gas has been strongly perturbed within a time as recent as a few orbital periods ($\sim 10^5$ yr). This disturbance was possibly produced by material which fell into

the galactic centre from an appreciable distance. Estimates from the amount of material present and its clumpy distribution indicate that material should fall into the centre at an average rate of $\sim 10^{-4}-10^{-3} M_\odot \text{ yr}^{-1}$ (ref. 14). This average rate should generate substantially more energy flux than is presently seen in the region. However, if there were episodic infall, the present perturbations and possibly also the low density of material within the inner radius of the disk could be the result of a past larger infall and very high resultant energy flux. This would also imply that we are now in a time of relative quiescence.

It has previously been argued that the ultraviolet radiation required to ionize gas in the galactic centre could not come from a single point-source, because a black-body source or one with the characteristics of a stellar atmosphere would not produce enough ultraviolet without being a stronger source of infrared than has been observed^{27,28}. Now we find that these arguments are not conclusive. Henry *et al.*¹⁰ have already suggested that much of the radiation could originate from the source IRS16. It does seem possible to fit the amount of ultraviolet radiation required to ionize the Sgr A West H II region and the observed infrared intensity of IRS16 to what would be expected from a radiating atmosphere similar to that of an O-type star of surface temperature $\sim 35,000$ K and surface gravity of $\log g = 5$ (ref. 1), although a large number of O-type stars would be required. The surface gravity tends to decrease radiation energetic enough to ionize Ar⁺ or S²⁺, fulfilling the observed conditions of very few Ar²⁺ and S³⁺ ions, while at the same time providing enough photons to produce the observed Ne⁺. Thus ultraviolet and infrared observations seem to be consistent with production of the ultraviolet excitation in the central few parsecs by a single localized source, such as a black hole.

Conclusions

New infrared and submillimetre measurements of the dynamics and distribution of material around the galactic centre have increased knowledge of the structure and mass distribution of this region. A large amount of data seem to fit a model, involving: (1) an unusual concentration of mass not in the form of ordinary stars, most probably a central black hole of mass $4 \pm 1 \times 10^6 M_\odot$; (2) an ionized cavity of relatively low gas density surrounding the centre and of radius 1.7 pc. This cavity also contains an additional $1-2 \times 10^6 M_\odot$ mass in a stellar cluster; (3) a circulating disk of dense ($\sim 10^5 \text{ cm}^{-3}$), warm (300 K) and clumpy neutral gas between 1.7 and ~ 10 pc which is ionized along much of its inner edge. Irregular gas velocities indicate strong perturbations of the motions in this region over the past few 10^5 yr.

The observed velocities also fit a stellar cluster with no central point mass if the stellar density distribution decreases at least as fast as $R^{-2.7}$. However, the case without a central black hole is excluded if the observed 2- μm light distribution is a fair representation of the mass distribution of the stellar cluster.

We thank M. Silber and D. M. Watson for help in obtaining these results. We also thank D. Backer, I. King, C. F. McKee and F. Shu for helpful discussions. This work was supported by NASA grants NAG 2-208 and NGL 05-003-272, NSF grants AST 8211520 and AST 8212055 and a Research Corporation Grant to R.G.

Received 29 January; accepted 18 April 1985.

- Serabyn, E. & Lacy, J. H. *Bull. Am. astr. Soc.* **16**, 980 (1984); *Astrophys. J.* (in the press).
- Genzel, R., Watson, D. M., Crawford, M. K. & Townes, C. H. *Astrophys. J.* (submitted).
- Lugten, J. B., Crawford, M. K., Genzel, R. & Townes, C. H. *Astrophys. J.* (submitted).
- Harris, A., Jaffe, D. T., Silber, M. & Genzel, R. *Bull. Am. astr. Soc.* **16**, 936 (1984); *Astrophys. J. Lett.* (submitted).
- Ekers, R. D., van Gorkom, J. H., Schwarz, U. J. & Goss, W. M. *Astr. Astrophys.* **122**, 143-150 (1983).
- Brown, R. L., Johnston, K. J. & Lo, K. Y. *Astrophys. J.* **250**, 155-159 (1981).
- Lo, K. Y. & Claussen, M. J. *Nature* **306**, 647-651 (1983).
- Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265-270 (1974).
- Lo, K. Y. in *The Galactic Center* (eds Riegler, G. & Blandford, R. D.) (Am. Inst. Phys., New York, 1974).
- Henry, J. P., DePoy, D. L. & Becklin, E. E. *Astrophys. J.* **285**, L27-L30 (1984).
- Storey, J. W. V. & Allen, D. A. *Mon. Not. R. astr. Soc.* **204**, 1153-1161 (1984).
- Hall, D. N. B., Kleinmann, S. G. & Scoville, N. Z. *Astrophys. J. Lett.* **260**, L53-L57 (1982).
- Geballe, T. R. *et al. Astrophys. J.* **284**, 118-125 (1984).

- Lacy, J. H., Townes, C. H., Geballe, T. R. & Hollenbach, D. J. *Astrophys. J.* **241**, 132-146 (1980).
- van Gorkom, J. H., Schwarz, U. J. & Bregman, J. D. *IAU Symp.* No. 106 (1984).
- Becklin, E. E., Gatley, I. & Werner, M. W. *Astrophys. J.* **258**, 135-142 (1982).
- Genzel, R. *et al. Astrophys. J.* **276**, 551-559 (1984).
- Gatley, I., Jones, T. J., Hyland, A. R., Beattie, D. H. & Lee, T. J. *Mon. Not. R. astr. Soc.* **210**, 565-575 (1984).
- Liszt, H. S., van de Hulst, J. M., Burton, W. B. & Ondrechen, M. P. *Astr. Astrophys.* **126**, 341-351 (1983).
- Serabyn, E. thesis, Univ. Calif., Berkeley (1984).
- King, I. R. *Q. Jl R. astr. Soc.* **22**, 227-243 (1981).
- Newell, B., DaCosta, G. S. & Norris, J. *Astrophys. J. Lett.* **208**, L55-L59 (1976).
- Becklin, E. E. & Neugebauer, G. *Astrophys. J.* **151**, 145-161 (1968).
- Allen, D. A., Hyland, A. R. & Jones, T. J. *Mon. Not. R. astr. Soc.* **204**, 1145-1152 (1984).
- Bahcall, J. N. & Wolf, R. A. *Astrophys. J.* **209**, 214-232 (1976).
- Quinn, P. J. & Sussman, G. J. preprint, Cal. Inst. Technology (1984).
- Lacy, J. H., Townes, C. H. & Hollenbach, D. J. *Astrophys. J.* **262**, 120-134 (1982).
- Townes, C. H., Lacy, J. H., Geballe, T. R. & Hollenbach, D. J. *Nature* **301**, 661-666 (1983).

Terrestrial xenon isotope constraints on the early history of the Earth

M. Ozima*, F. A. Podosek† & G. Igarashi*

* Geophysical Institute, University of Tokyo, Tokyo 113, Japan

† Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA

Comparison between ^{129}I -radiogenic ^{129}Xe and ^{244}Pu -fissiogenic ^{136}Xe components in terrestrial xenon suggests that the Earth's inner region accreted a few tens of millions of years earlier than the outer region from which the atmosphere evolved. The results also indicate that there has been no substantial mixing of the two regions since the Earth's accretion.

SINCE the discovery of excess ^{129}Xe in meteorites¹, Xe isotopes produced from ^{129}I (half-life $t_{1/2} = 17$ Myr) and from ^{244}Pu ($t_{1/2} = 82$ Myr) have been a prime source of information on the early history of meteorites. Although there has been considerable work on Xe in meteorites, there is still relatively little Xe isotope data for terrestrial samples: nevertheless, some very interesting results have emerged, notably the confirmation of large excess ^{129}Xe and $^{132-136}\text{Xe}$, relative to air Xe, in gas from a CO_2 well in New Mexico²⁻⁵. There have also been reports of excess ^{129}Xe in some mantle-derived materials, such as mid-ocean ridge basalts (MORB)⁶⁻⁸, mantle xenoliths⁹⁻¹¹ and diamonds¹². The existence of excess ^{129}Xe indicates that the source region for the excess ^{129}Xe must have been isolated for almost the whole history of the Earth from the atmosphere and/or the rest of the mantle from which the atmosphere was generated. It has been argued^{6,7,13-15} that the isolation of this source must correspond to degassing of the mantle and, therefore, that this degassing, and the major generation of the atmosphere, occurred very early, within a few half-lives of ^{129}I after the formation of the Earth.

We now examine published data on Xe isotope composition in terrestrial samples and discuss their implications for the early history of the Earth. We show that the terrestrial Xe isotope data need not lead to the conclusion that the excess ^{129}Xe requires very early mantle degassing, as generally assumed.

Terrestrial Xe isotope data

From published Xe data, we have constructed three isotope plots for granites, MORBs and upper-mantle derived rocks (mostly mantle xenoliths) (Figs 1-3). The data for New Mexico CO_2 -well gas Xe is shown in the plots for granites (Fig. 1). In plotting these correlation diagrams our only selection criterion was to discard those with abnormally large error assignments. For step-heating data, we discarded the lowest temperature fraction, which would probably include large air contamination, and we chose as a representative value, the fraction with the largest amount of Xe. For the New Mexico CO_2 -well Xe, we used the latest experimental data⁵ (Fig. 1).

Granites (Fig. 1) show a very clear linear correlation which must be due to the addition of ^{238}U -fissiogenic Xe isotopes¹⁶ as granite typically has relatively high concentrations of U. The slopes of the correlation lines (solid lines) in the $^{134}\text{Xe}/^{130}\text{Xe}$ - $^{136}\text{Xe}/^{130}\text{Xe}$ and $^{132}\text{Xe}/^{130}\text{Xe}$ - $^{136}\text{Xe}/^{130}\text{Xe}$ diagrams give the ratios of the fission yields $^{134}\text{Y}/^{136}\text{Y}$ and $^{132}\text{Y}/^{136}\text{Y}$, respectively, which agree well with the laboratory values of 0.832 ± 0.02 and 0.59 ± 0.017 for spontaneous fission of ^{238}U . Figure 1 also shows the correlation lines (dotted lines) corresponding to ^{244}Pu -fissiogenic Xe, which do not fit the data well. The isotope composition for New Mexico CO_2 -well Xe falls on the correlation lines, indicating that the excess ^{132}Xe - ^{136}Xe is also due to ^{238}U -fission, as was concluded by Phinney *et al.*⁴. However, the CO_2 -well gas Xe has significant excess ^{129}Xe , which is not seen in granites. This excess ^{129}Xe has been confirmed independently³⁻⁵ and its existence seems to be established.

Excess ^{129}Xe may also exist in MORB (Fig. 2) and mantle-derived xenoliths (Fig. 3). In addition, ^{136}Xe and ^{134}Xe appear

to show slight excesses relative to air Xe. However, correlations between $^{136}\text{Xe}/^{130}\text{Xe}$ and $^{134}\text{Xe}/^{130}\text{Xe}$, $^{132}\text{Xe}/^{130}\text{Xe}$ are vague and it is not possible to resolve whether the excess $^{132-136}\text{Xe}$, if real, is due to ^{238}U -fission or to ^{244}Pu -fission. The only known source for excess ^{129}Xe is the extinct nuclide ^{129}I . The existence of the excess ^{129}Xe indicates that some part of the Earth had incorporated a non-trivial amount of short-lived ^{129}I and has not undergone extensive exchange with the atmosphere, or the source of the atmosphere, since before the decay of ^{129}I was substantially complete. For excess ^{131}Xe - ^{136}Xe , both ^{238}U -fission and ^{244}Pu -fission could be responsible. Except for crustal rocks, which are more enriched in U, ^{244}Pu -fissiogenic Xe is likely to be more important than ^{238}U -fissiogenic Xe in the Earth as a whole¹⁷.

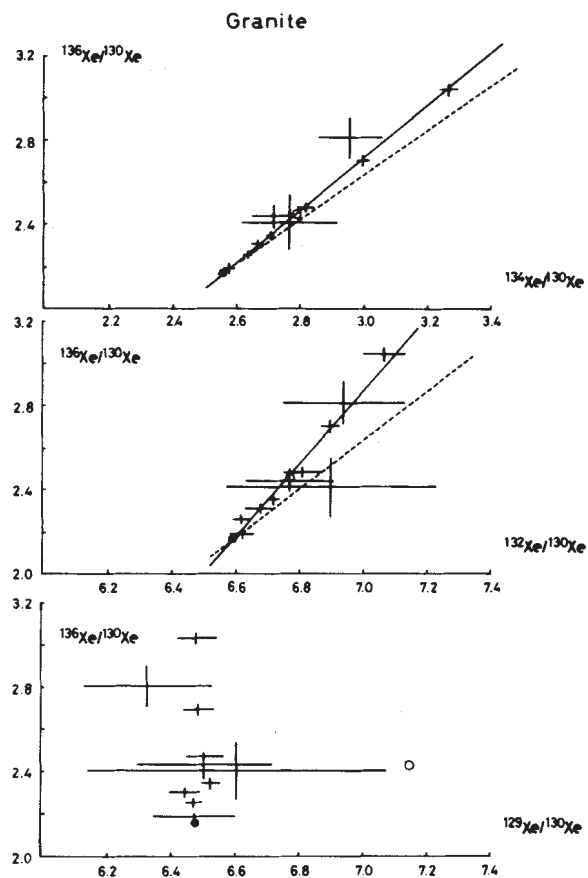


Fig. 1 Three isotope plots for granite samples¹⁶. Also shown are New Mexico CO_2 -well gas Xe (○, data from ref. 5) and air Xe (●). Note that the CO_2 -well gas Xe data lie on the correlation lines, which characterize ^{238}U -fissiogenic Xe (solid lines). For comparison, a correlation line for ^{244}Pu -fissiogenic Xe is also shown by a dotted line.

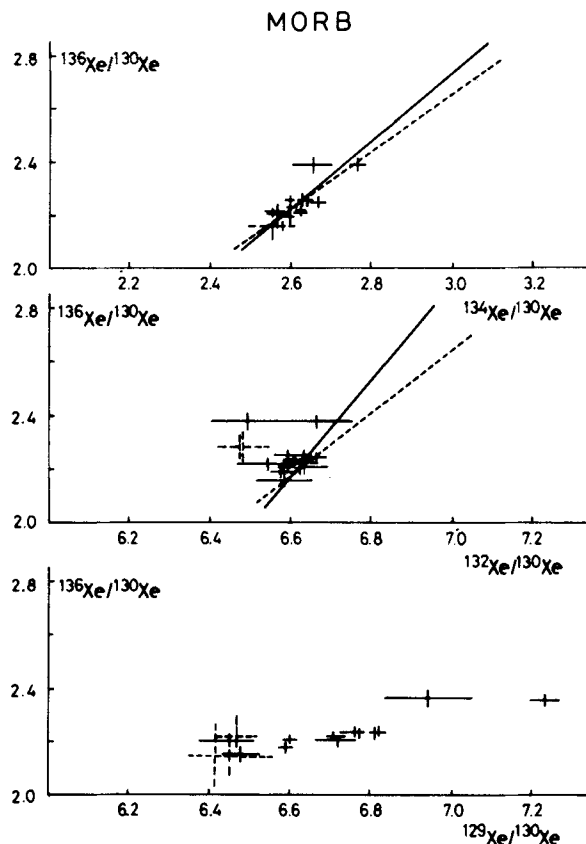


Fig. 2 Three isotope plots for MORB. Solid cross, data from refs 6, 7. Dotted cross, data from ref. 27. Note that neither the ^{244}Pu -fissionogenic nor the ^{238}U -fissionogenic lines fit the data.

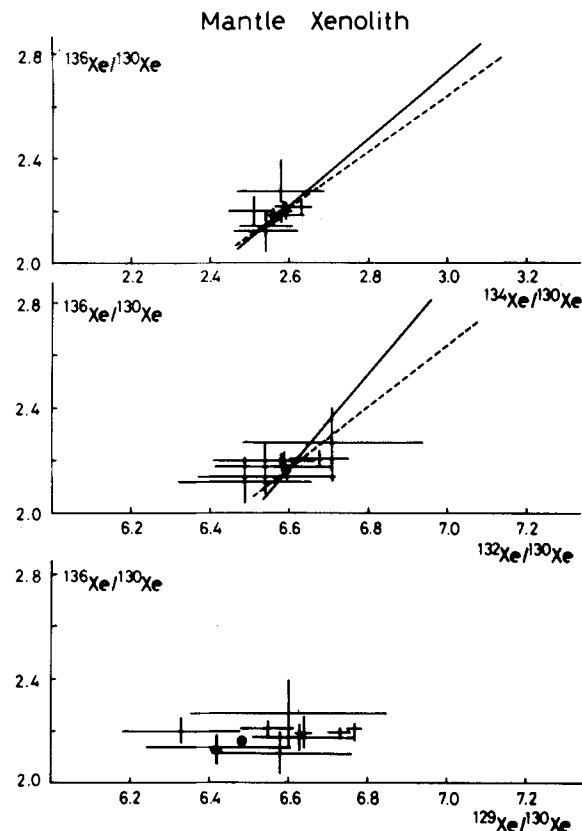


Fig. 3 Three isotope plots for mantle-derived xenoliths (data from refs 9, 10, 28). ●, Air Xe. A slight excess may be seen in ^{129}Xe .

Xe closure time

In using radiogenic isotopes as chronological tracers, a fundamental problem is to identify when the parent-daughter element system closed. The nucleogenic Xe isotopes that are important in terrestrial Xe are spontaneous fissionogenic Xe (^{131}Xe , ^{132}Xe , ^{134}Xe , ^{136}Xe) from ^{238}U and ^{244}Pu and radiogenic ^{129}Xe . Once the system is closed for $^{129}\text{I}/\text{Xe}$ and $^{244}\text{Pu}/\text{Xe}$ ratios, Xe isotope evolution obeys a simple radioactive decay law and is predictable. Because our main concern here is the early history of the Earth and thus with I-Xe and Pu-Xe closure; we can disregard the ^{238}U -Xe system in much of the following discussion because of the long half-life of ^{238}U .

In gas of a solar composition, Xe isotope evolution would be negligible because of very low I/Xe and Pu/Xe ratios; only in condensed materials greatly depleted in noble gases would Xe evolution be significant. Theoretical treatments of early Solar System history typically conclude that condensation and fragmentation into planetesimals (bodies a few kilometres in diameter) would occur quite rapidly, on a 10^4 -yr timescale at the Earth's orbit^{18,19}, but that accretion of planetesimals into the major terrestrial planets might take significantly longer, of the order of 10^7 - 10^8 yr (ref. 20). This is a significant time for short-lived ^{129}I and ^{244}Pu . While condensation and accretion result in closure of solid elements such as Pu and I, Xe closure might be established only after collisions among planetesimals, because impact shock could result in the effective degassing of volatiles²¹. In this way, it would be possible for early and late-accreting material to have significantly different Xe closure times and contribute significantly different amounts of radiogenic Xe, especially ^{129}Xe , even if it all had the same basic chemistry and the same initial I/Pu ratio.

Interpretation

According to Pepin and Phinney^{22,23}, about 6.7% of atmospheric ^{129}Xe originates from the decay of ^{129}I , and 4.7% of atmospheric

^{136}Xe originates from the decay of ^{244}Pu . Subtracting these nucleogenic components from the air Xe composition, we can define a 'primary' air Xe (equivalent to "corrected atmosphere Xe"²³); this is presumably the Xe composition at the time of isotopic closure. It is a straightforward matter to develop equations for the compositional evolution after closure and to include a single episodic degassing event. Using subscripts '0' for closure and 'd' for degassing, and setting $t_0 = 0$, we have for $0 \leq t \leq t_d$ (from Xe closure to degassing):

$$\begin{aligned} \frac{^{129}\text{Xe}}{^{130}\text{Xe}} &= \left(\frac{^{129}\text{Xe}}{^{130}\text{Xe}} \right)_0 + \left(\frac{^{129}\text{I}}{^{130}\text{Xe}} \right)_0 (1 - \exp(-\lambda_{129}t)) \\ \frac{^{136}\text{Xe}}{^{130}\text{Xe}} &= \left(\frac{^{136}\text{Xe}}{^{130}\text{Xe}} \right)_0 + \left(\frac{^{244}\text{Pu}}{^{130}\text{Xe}} \right)_0 {}^{136}\text{Y}_{\text{Pu}} (1 - \exp(-\lambda_{244}t)) \end{aligned} \quad (1)$$

where λ_{129} and λ_{244} are the decay constants for ^{129}I and ^{244}Pu and ${}^{136}\text{Y}_{\text{Pu}}$ is the yield of ^{136}Xe per decay of ^{244}Pu .

If the degassing occurred before ^{129}I and ^{244}Pu had decayed completely, Xe isotopic composition in the residue continued to evolve. If we assume that a fraction $(1-f)$ of the Xe was degassed at $t = t_d$, the evolution for $t > t_d$ becomes

$$\begin{aligned} \frac{^{129}\text{Xe}}{^{130}\text{Xe}} &= \left(\frac{^{129}\text{Xe}}{^{130}\text{Xe}} \right)_0 + \left(\frac{^{129}\text{I}}{^{130}\text{Xe}} \right)_0 (1 - \exp(-\lambda_{129}t_d)) \\ &\quad + \frac{1}{f} \left(\frac{^{129}\text{I}}{^{130}\text{Xe}} \right)_0 \exp(-\lambda_{129}t_d) \\ &\quad \times (1 - \exp(-\lambda_{129}(t - t_d))) \\ \frac{^{136}\text{Xe}}{^{130}\text{Xe}} &= \left(\frac{^{136}\text{Xe}}{^{130}\text{Xe}} \right)_0 + \left(\frac{^{244}\text{Pu}}{^{130}\text{Xe}} \right)_0 {}^{136}\text{Y}_{\text{Pu}} (1 - \exp(-\lambda_{244}t_d)) \\ &\quad + \frac{1}{f} \left(\frac{^{244}\text{Pu}}{^{130}\text{Xe}} \right)_0 {}^{136}\text{Y}_{\text{Pu}} \exp(-\lambda_{244}t_d) \\ &\quad \times (1 - \exp(-\lambda_{244}(t - t_d))). \end{aligned} \quad (2)$$

Figure 4 shows a single-stage Xe-isotope evolution curve, for

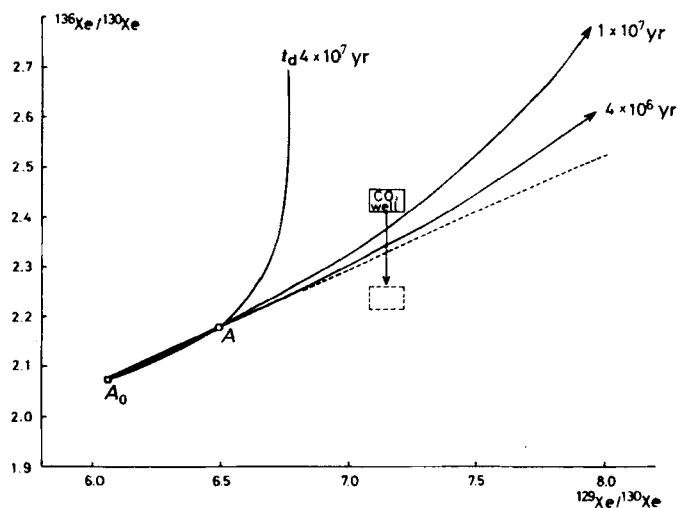


Fig. 4 Xe isotope evolution model in the Earth. A, Air Xe; A_0 , primary air Xe (see text). A_0 is the Xe-closure time in the Earth. Owing to ^{129}I and ^{244}Pu , Xe isotopes in the Earth had evolved along the paths A_0 –A. At t_d Xe was degassed into the atmosphere, where the Xe isotopic ratio was quenched to give air Xe (A). After t_d , Xe in the Earth has continued to evolve. The evolution curve is determined by specifying t_d . We also show the CO_2 -well gas Xe (refs 4, 5). Solid rectangle, the range for the measured isotope composition; dotted rectangle, isotopic composition (maximum value) corrected for the ^{238}U -fissionogenic ^{136}Xe .

which atmospheric Xe is assumed to be degassed at $t = t_d$. Although drawn for $f = 0.5$, the choice of f ($0 < f < 1$) does not significantly affect the general character of the evolution. Equation (1) clearly shows that the specification of t_d determines $(^{244}\text{Pu}/^{130}\text{Xe})_0$ and $(^{129}\text{I}/^{130}\text{Xe})_0$, as both $(^{136}\text{Xe}/^{130}\text{Xe})_{\text{air}}$ and $(^{129}\text{Xe}/^{130}\text{Xe})_{\text{air}}$ are assumed to be known. This situation can also be seen in Fig. 4: an evolution curve passing through A_0 (primary air Xe) and A (present air Xe) can be uniquely determined by specifying t_d . The shorter the t_d , the further the curves grow after the degassing, since shorter t_d requires greater I/Xe and Pu/Xe and much of the decay occurs after the degassing. Note that Xe isotope composition in any degassed region ($t_d > 0$) must be located in the region above the dotted line, that is, the extension of the line connecting A_0 and A in Fig. 4; qualitatively, the ratio of fission ^{136}Xe to radiogenic ^{129}Xe must increase with time because of the greater half-life of ^{244}Pu .

Let us now consider the CO_2 -well gas Xe. Figure 4 plots the CO_2 -well gas Xe both as measured and as corrected for ^{238}U fissionogenic ^{136}Xe . Phinney *et al.*⁴ estimated that at least 80% (and perhaps all) of the excess ^{136}Xe is due to ^{238}U -fission (rather than ^{244}Pu fission). We have re-estimated the fraction of the ^{238}U -contribution to the excess ^{136}Xe using a least-squares method, with the result that it is at least 80%. Figure 4 shows that the CO_2 -well gas Xe corrected for ^{238}U -fission lies below the dotted line, where the correction involves subtracting 80% of the excess $^{136}\text{Xe}/^{130}\text{Xe}$, relative to the air ratio, from the observed ratio (that is $(^{136}\text{Xe}/^{130}\text{Xe})_{\text{corr.}} = 2.448 - (2.448 - 2.1747) \times 0.8 = 2.2294$). The corrected value should be an upper limit, as the ^{238}U -fissionogenic contribution is likely to be larger than 80%. By the above argument, this observation indicates that the CO_2 -well gas Xe (corrected for ^{238}U fission) was not derived from the residue from which atmospheric Xe was degassed, nor entirely as a mixture of air with Xe from this residue, because such Xe must lie above the dotted line.

The fundamental difference between the source region of the well-gas Xe and that of air Xe is that the former must have had a smaller $(^{244}\text{Pu}/^{129}\text{I})_0$ value. It is easy to show that the Xe isotope evolution curve which evolved from A_0 to terminate at the nominally corrected well-gas composition is uniquely specified by a $(^{244}\text{Pu}/^{129}\text{I})_0$ value of 2,105, which is much smaller than the minimum permissible value for the source of atmospheric Xe ($(^{244}\text{Pu}/^{129}\text{I})_{\text{min}} = 3,306$, corresponding for $t_d \gg 1/\lambda_{244}$). This difference may have resulted from chemical differ-

entiation in the Earth or from chemical heterogeneity among the planetesimals which accreted to form the Earth. In the former case, the chemical differentiation must have occurred before ^{129}I or ^{244}Pu had decayed, that is, within a few tens of millions of years after the formation of the Earth. In the latter case, the difference in $(^{244}\text{Pu}/^{129}\text{I})_0$ among planetesimals may reflect a variation in the Xe-closure time of different planetesimals. If the time lag is responsible for the difference in the $(^{244}\text{Pu}/^{129}\text{I})_0$ ratio, the time interval required can be estimated from

$$\left(\frac{^{244}\text{Pu}}{^{129}\text{I}}\right)_{0,\text{outer}} / \left(\frac{^{244}\text{Pu}}{^{129}\text{I}}\right)_{0,\text{inner}} = \exp\{(\lambda_{129} - \lambda_{244})\Delta t\} \quad (3)$$

Setting $(^{244}\text{Pu}/^{129}\text{I})_{\text{outer}} > 3,306$ and $(^{244}\text{Pu}/^{129}\text{I})_{\text{inner}} = 2,105$ in the left-hand side of equation (3). We obtain $\Delta t > 1.4 \times 10^7$ yr.

Earth's early history

The excess ^{129}Xe in some terrestrial materials indicates that the region from which the excess ^{129}Xe was derived must have been isolated from the atmosphere for nearly the whole history of the Earth. We argued above against the commonly held view^{6,7,13–15} that the source region for the excess ^{129}Xe in the CO_2 -well gas represents the degassed mantle from which air Xe was derived.

We propose the following scenario for interpreting coherently the available Xe isotope data. The Earth formed by accretion of impacting planetesimals. Planetesimals were originally formed by the fragmentation process²⁴, underwent mutual collision to grow further and finally impacted onto the Earth. The early planetesimals constituted the inner part of the Earth. As the accretion time of the Earth is plausibly at least a few tens of millions of years^{18,19} and the time of the final impact of the planetesimals onto the Earth must be an upper limit to their Xe-closure time, there could well have been a few tens of millions of years time gap in Xe-closure time between the inner and outer regions of the Earth. Hence, we expect that the inner part of the Earth had smaller $(^{244}\text{Pu}/^{129}\text{I})_0$ as well as larger $(^{244}\text{Pu}/^{130}\text{Xe})_0$ and $(^{129}\text{I}/^{130}\text{Xe})_0$ ratios. If we assume that the CO_2 -well gas Xe (except for the ^{238}U fission component) was derived from the inner part of the Earth, while air Xe was degassed from the outer region, we could explain why the CO_2 -well gas Xe is characterized by excess ^{129}Xe but not a corresponding excess of ^{244}Pu -derived $^{131–136}\text{Xe}$. The comparison between air Xe and CO_2 -well gas Xe then gives about 1.4×10^7 yr for the minimum time interval between the formation of the two regions. If the timescale for atmospheric degassing is of the order of 10^8 yr, as inferred from very high $^{40}\text{Ar}/^{36}\text{Ar}$ ratios in the upper mantle²⁵, then the upper-mantle Xe composition should be nearly the same as the atmospheric Xe composition.

Excess ^{129}Xe has been reported in some mantle-derived samples. The largest excess ^{129}Xe reported for terrestrial materials is for MORB by Allègre *et al.*⁷ who argued that the excess ^{129}Xe in MORB is indicative of very early degassing of the MORB source region. The argument can be examined by means of Fig. 4. However, the difficulty in identifying ^{244}Pu -fissionogenic Xe does not allow such a test. Hence, until the origin of the excess $^{131–136}\text{Xe}$ is resolved, the conclusion by Allègre *et al.* should be treated with caution.

The high $^{40}\text{Ar}/^{36}\text{Ar}$ ratios in MORBs relative to air Ar are not inconsistent with the rapid timescale for degassing (within a few tens of millions years after Earth's formation) argued by several authors on the basis of the excess ^{129}Xe in the mantle-derived materials^{6,7,13–15}. However, note that while the high $^{40}\text{Ar}/^{36}\text{Ar}$ ratio has been established throughout the Earth's history owing to the much slower decay of ^{40}K , the excess ^{129}Xe was established soon after the formation of the Earth. In the upper mantle, where material movement such as mantle convection and subduction are likely to have operated throughout geological time, it is difficult to suppose that the early established excess ^{129}Xe could have survived these geological disturbances. The mixing of subducted materials into the mantle is especially difficult to reconcile with the persistent existence of excess Xe

in MORB because, unlike lighter noble gases, Xe is notoriously 'sticky' and is likely to be carried down into the mantle with the subducting ocean sediments, which have 2-4 orders of magnitude enrichment of Xe relative to common volcanic rocks²⁶. Hence, it may be more reasonable to suppose that the excess ¹²⁹Xe observed in MORBs (and also in other mantle-

derived materials) did not originate in the degassed upper mantle, but was transported there from some other region (possibly from the deeper mantle). This source had been isolated from the mantle-atmosphere system for almost the whole history of the Earth as shown by the presence of excess ¹²⁹Xe in the CO₂-well gas Xe.

Received 16 October 1984; accepted 28 March 1985.

1. Reynolds, J. H. *Phys. Rev. Lett.* **4**, 8-10 (1960).
2. Butler, W. A., Jeffrey, P. M., Reynolds, J. H. & Wasserburg, G. J. *J. geophys. Res.* **68**, 3283-3291 (1963).
3. Boulou, M. S. & Manuel, O. K. *Science* **174**, 1334-1336 (1971).
4. Phinney, D., Tennyson, J. & Frick, U. *J. geophys. Res.* **83**, 2313-2319 (1978).
5. Smith, S. P. & Reynolds, J. H. *Earth planet. Sci. Lett.* **54**, 236-238 (1981).
6. Staudacher, Th. & Allègre, C. J. *Earth planet. Sci. Lett.* **60**, 389-406 (1982).
7. Allègre, C. J., Staudacher, Th., Sarda, P. & Kurz, M. *Nature* **303**, 762-766 (1983).
8. Kirsten, T., Oehm, J., Richter, H. & Rocholl, A. *Meteoritics* **18**, 355 (1983).
9. Kaneoka, I. & Takaoka, N. *Earth planet. Sci. Lett.* **39**, 382-386 (1978a).
10. Kaneoka, I., Takaoka, N. & Aoki, K. in *Terrestrial Rare Gases* (eds Alexander, E. C. & Ozima, M.), 71-83 (Japan Scientific Society Press, Tokyo, 1978b).
11. Henneke, E. W. & Manuel, O. K. *Nature* **257**, 778-780 (1975).
12. Takaoka, N. & Ozima, M. *Nature* **271**, 45-46 (1978).
13. Manuel, O. K. & Sabu, D. D. *Geochem. J.* **15**, 245-267 (1981).
14. Thomsen, K. *J. geophys. Res.* **85**, 4374-4378 (1980).
15. Hart, R., Dymond, J. & Hogan, I. *Nature* **305**, 403-407 (1983).
16. Kuroda, P. K., Sherrill, R. D. & Jackson, K. C. *Geochem. J.* **11**, 75-90 (1977).
17. Bernatowicz, T. J. & Podosek, F. A. in *Terrestrial Rare Gases* (eds Alexander, E. C. & Ozima, M.) 99-135 (Japan Scientific Society Press, Tokyo, 1978).
18. Nakagawa, Y., Nakazawa, K. & Hayashi, C. *Icarus* **45**, 517-528 (1981).
19. Weidenschilling, S. J. *Icarus* **44**, 172-189 (1980).
20. Nakagawa, Y., Hayashi, C. & Nakazawa, K. *Icarus* **54**, 361-376 (1983).
21. Lange, M. A. & Ahrens, T. J. *J. geophys. Res. Suppl.* **87**, A451-A456 (1982).
22. Pepin, R. O. & Phinney, D. *Lunar Sci.* **7**, 682-684 (1976).
23. Pepin, R. O. & Phinney, D. *Moon Planets* (in the press).
24. Goldreich, P. & Ward, W. R. *Astrophys. J.* **183**, 1051-1061 (1973).
25. Hamano, Y. & Ozima, M., in *Terrestrial Rare Gases* (eds Alexander, E. C. & Ozima, M.) (Japan Scientific Society Press, Tokyo, 1978).
26. Podosek, F. A., Honda, M. & Ozima, M. *Geochim. cosmochim. Acta* **44**, 1875-1884 (1980).
27. Ozima, M. & Zashu, S. *Earth planet. Sci. Lett.* **62**, 24-40 (1983).
28. Kyser, T. K. & Rison, W. J. *geophys. Res.* **87**, 5611-5630 (1982).

Quaternary structure of the acetylcholine receptor

A. Brisson & P. N. T. Unwin

Service de Physique, Groupe Structures, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires, Grenoble 38041, France and Department of Cell Biology, Stanford University School of Medicine, Stanford, California 94305, USA

The five membrane-spanning subunits of the acetylcholine receptor have been resolved in electron microscope images and are shown to lie at pentagonally symmetrical positions around the channel over a large fraction of their length. The channel consists of a wide synaptic portion and a narrow portion extending through the membrane into the interior of the cell.

SIGNALLING between excitable cells is commonly mediated by the release of neurotransmitters, which bind subsequently to receptors in the target membranes and induce in them a change of permeability. The nicotinic acetylcholine receptor, the best understood such receptor, effects the change by opening a channel in the membrane through which cations, principally sodium and potassium, diffuse. This action is initiated by the cooperative binding of neurotransmitter to the α -subunits, the two smallest of five (stoichiometry $\alpha_2\beta\gamma\delta$) assembled in a ring around the channel (see refs 1-3 for review).

The complete amino-acid sequences of the four types of polypeptide comprising this receptor have recently been deduced⁴⁻⁷ and indicate extensive homology between the subunits (see ref. 8 for review). Predictions based on these data suggest a common motif of secondary structure for the portions

of the subunits inside the lipid bilayer^{9,10}, with features having important implications concerning the movement of ions through the channel and the mechanism of the gating event^{11,12}. An understanding of the molecular details now requires direct structural information on the different transitional states.

Here we report on a quaternary configuration of the acetylcholine receptor, determined by three-dimensional electron image analysis of tubular crystals grown from native membrane vesicles. Previously, a pentagonal molecular outline had been resolved in individual images of stained tubes, suggesting that the subunits are disposed regularly around the channel¹³. The three-dimensional maps obtained by combining information from many images now reveal details of the subunits and of the channel at different levels throughout the structure. We demonstrate that there is indeed an approximate 5-fold axis of symmetry along the channel axis over a large fraction of the molecule extending from within the lipid bilayer into the synaptic cleft.

Table 1 Three-dimensional data from images of tubes

	Frozen	Stained
No. of images*	35	52
Range of tilt angles	0-60°	0-65°
No. of independent lattice lines†	21	24
No. of Fourier terms used in synthesis‡	208	271
Average phase error/image§	13.5°	11.9°

* Values for the axial ratio b/a and included angle γ were 1.787 ± 0.01 , $120.17 \pm 0.19^\circ$ (frozen) and 1.797 ± 0.01 , $117.33 \pm 0.48^\circ$ (stained); from 20 measurements, with a , b and γ as defined previously¹³. Thus, the ratio of lateral to longitudinal dimensions is ~5% smaller for the stained than for the frozen tubes.

† Data include all significant lattice lines (except for the (0, 0)) to a resolution of 25 Å; these are the (0, 2), (0, 3), (0, 4), (0, 5), (1, -5), (1, -4), (1, -3), (1, -2), (1, -1), (1, 0), (1, 1), (1, 2), (1, 3), (2, -5), (2, -4), (2, -3), (2, -2), (2, -1), (2, 0), (3, -3), (3, -1) lattice lines in both cases, and in addition the (0, 1), (3, -5), (3, -2) for the tubes in stain.

‡ Obtained by sampling the continuous amplitude and phase curves at regular 0.0025-Å⁻¹ intervals.

§ Based on cumulative comparisons of individual peaks over a range of 0.003 Å⁻¹ along the lattice lines.

Data collection and analysis

The tubes, which were prepared from the electric organ of the ray *Torpedo marmorata*, are composed of receptors arranged in a p2 surface lattice and oriented with the synaptic side outwards¹³. They are flattened against the support film during preparation (see Fig. 2 legend), forming two almost planar superposed layers (Fig. 1). Images of the tubes in frozen solution and in negative stain were obtained to provide information on the complete protein and the hydrophilic surfaces, respectively.

For the analysis, all the selected tubes had surface lattices in orientations close to the example in Fig. 1b; therefore, variations in their unit cell parameters were less than described previously¹³. The tubes were tilted at angles of up to 65° to the incident electron beam and the amplitudes and phases of the Fourier components were determined to a resolution of 25 Å from digitized areas of these images. Data were then combined cumulatively by standard methods¹⁴ to map the continuous variations of amplitude and phase along each reciprocal lattice line (Fig. 2). The three-dimensional Fourier syntheses were from the terms obtained by sampling these curves at regular intervals

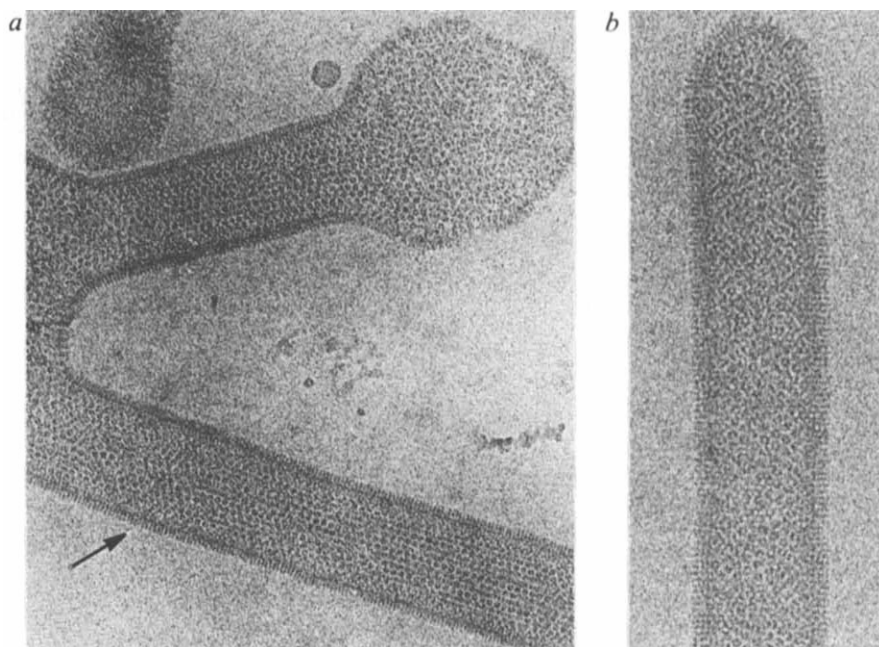


Fig. 1 Electron microscope images of tubular crystals of acetylcholine receptor embedded in thin films of frozen buffer (100 mM Tris-HCl, pH 6.8). The complex patterns arise from superposition of *en face* views of the receptors located on the top and bottom surfaces. Striated zones at the tube borders are a result of the receptors being viewed more nearly from the side; when aligned parallel to the tube axis, as in *a*, the receptors in these zones form distinct 120–140-Å wide bands (arrow) and seem to be viewed exactly from the side. Magnifications: *a*, $\times 112,000$; *b*, $\times 182,400$.

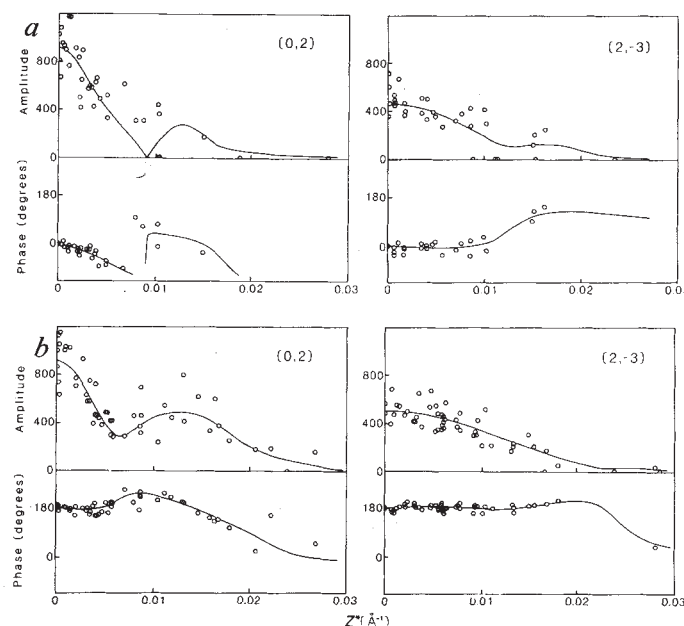
and assuming the unit cell dimensions in the plane of the membrane to be $a = 90.0$ Å, $b = 160.8$ Å, $\gamma = 120.2^\circ$, determined from the frozen specimens. The densities synthesized from the stained specimens were multiplied by -1 to correct for their opposite contrast. Further details are given in Table 1.

The similarity between the two three-dimensional maps was evaluated by comparing sections at various levels throughout the two structures. The sections were parallel to the plane of the membrane and the comparisons included all the densities

in a single unit cell. The best overall correlation was achieved with a relative displacement of 11 Å perpendicular to the membrane between the origins of the two maps, reflecting differences between the centres of scattering density for the two cases. Corresponding values for the correlation coefficients were >0.8 over most of the ~ 140 Å length in which features were clearly contrasted (Fig. 3), except for a narrow zone identified as the hydrophobic core of the bilayer (see below). Hence, with the exception of this zone, the two maps resemble each other closely.

Fig. 2 Comparison of amplitude and phase variations along typical lattice lines after combining images from tubes in: *a*, ice; *b*, stain, according to the 2-sided plane group p21. The continuous curves through the data points were computed by a least-squares method with amplitudes weighted according to the reliability of the corresponding phases²². Z^* is the distance from the midpoint of the lattice line.

Methods. Tubes were prepared as described previously¹³ and applied to carbon support grids rendered highly charged by glow discharge in amylamine vapour²³. For freezing, grids were blotted with filter paper for 4 s, then plunged rapidly into liquid ethane cooled to below -130°C ; for staining, they were washed in 2% sodium phosphotungstate pH 7.2 and then dried. Frozen specimens were stored under liquid nitrogen and later examined in a Philips EM400 electron microscope using a cold holder maintained at -174°C . The electron optical magnifications were $\times 40,800$ (ice) and $\times 28,500$ (stain), and images were recorded at 100 kV on Kodak SO163 film developed for 8 min in concentrated D19 developer. A Philips low-dose kit was used to minimize radiation damage and to ensure that all images incorporated in the three-dimensional datasets were of tubes which had received total doses of no more than ~ 15 electrons per Å². Optical adjustments were performed at a nominal magnification on the screen of $\times 90,000$ to minimize variability in focus level from one image to the next. Tube images were evaluated by optical diffraction (see ref. 13) for the extent of structural preservation and degree of flattening, and then evaluated further by computer. For the computer analysis, densitometry and Fourier transformation of the central portion of the tubes were as described previously¹³; criteria for selection of images for incorporation into the three-dimensional datasets included the strength of the diffraction maxima above background and the accuracy with which they fit a regular two-dimensional net. Also, areas of interest were considered unsuitable if the corresponding under-focus values (see below) were outside the range 8–16,000 Å. Variability in lattice parameters between tubes necessitated the use of special methods to derive accurate values of tilt angle and axis. All images were recorded in pairs. With the frozen specimens (*a*) the second image of each pair was of the carbon support film only, after intensifying the electron beam to sublimate the ice. Diffraction patterns from these images displayed a series of concentric rings²⁴ extending to spacings <10 Å. Defocus values could be measured locally with an accuracy of ± 250 Å over a wide range, based on matching such patterns against a calibrated set. Tilt parameters were computed from the defocus values measured at the four corners of each film. Typical errors, estimated by comparing values from the four alternative planes passing through the different sets of three points, were 2° in tilt angle and 4° in direct of tilt axis, slightly larger errors applying to smaller tilts ($<10^\circ$). With the stained specimens (*b*), the second image was recorded (with negligible in-between irradiation) at another angle of tilt, differing from the first by a pre-selected amount (typically 15°). Tilt parameters were computed from the changes in reciprocal lattice geometry between the two images. In this case the independent values obtained from the two sides of each tube could be compared, and the errors indicated were found to be of similar magnitude as in the first method.



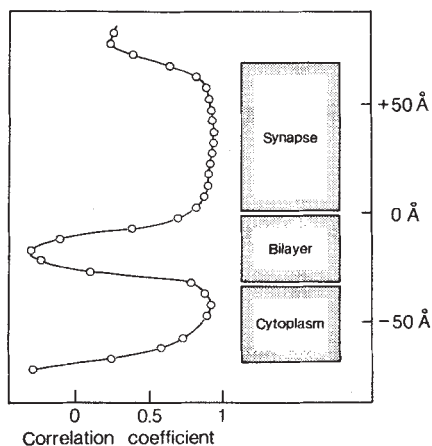


Fig. 3 Plot of correlation coefficients obtained by comparing sections from the three-dimensional maps of the frozen and stained tubes. The ordinate indicates the various levels through the structure, perpendicular to the membrane, at which the values were calculated. The hydrophobic core of the bilayer is identified by a zone showing poor correlation. The synaptic and cytoplasmic domains show good correlation over most of their extent; the poorer correlation near the extremities of the structure may be due to slight flattening of the stained receptors in the direction normal to the membrane plane.

Exact correspondence, particularly at the extremities of the structure, is not expected because of minor alterations resulting from, for example, drying of the stain¹⁵.

Structure of the receptor

The three-dimensional map calculated from the frozen tubes (Fig. 4) shows details of the entire receptor molecule, both in the lipid bilayer and in the aqueous surroundings on either side. The molecule is resolved into five rod-shaped subunits (Fig. 4a) of about equal cross-section, which lie predominantly perpendicular to the membrane plane. The five subunits are contained within a ~ 140 -Å-long by ~ 80 -Å-diameter cylindrical shell, and delineate a water-filled opening, presumed to be the channel, along the axis of this cylinder.

This map does not distinguish between the portion of the oligomer embedded in the bilayer and those protruding outside, because lipid and water contribute about equally to the image contrast at low resolution. The portions, however, can be identified by correlation with the map from the stained tubes. A ~ 30 -Å zone in that map appears almost featureless—neither the channel nor the surrounding protein is visible—and must correspond to the hydrophobic core of the bilayer, which is inaccessible to the stain. Consequently, there is a length of ~ 30 Å, the correlation coefficients of which have particularly low values (Fig. 3), and the matching of structural detail is poor (Fig. 5b) compared with that on either side (Fig. 5a, c). Thus the correlation places the hydrophobic core nearest to the cytoplasmic end of the structure, leaving a length of 35–45 Å protruding on this side, compared with ~ 70 Å on the other.

The channel is nearly 30 Å wide at the synaptic end of the oligomer (Fig. 4b, uppermost), but becomes too narrow at the cytoplasmic end to be seen. The way in which the channel narrows cannot be discerned accurately at the resolution achieved in the present study, but the abrupt exclusion of the anionic stain at the level of the bilayer surface, on the synaptic side, suggests that a large reduction in its diameter occurs in this region.

The cross-sectional area of the membrane-spanning portion of the subunits is of interest because of the different models that have been proposed for the internal structure of this portion. To estimate this value, we assumed the molecular boundary to be at the contour level for which the total enclosed volume of glycosylated protein is $350,400$ Å³, calculated from the composition⁸. The corresponding cross-sectional area per subunit was 540 Å², averaged over the extent of the bilayer, whereas for the

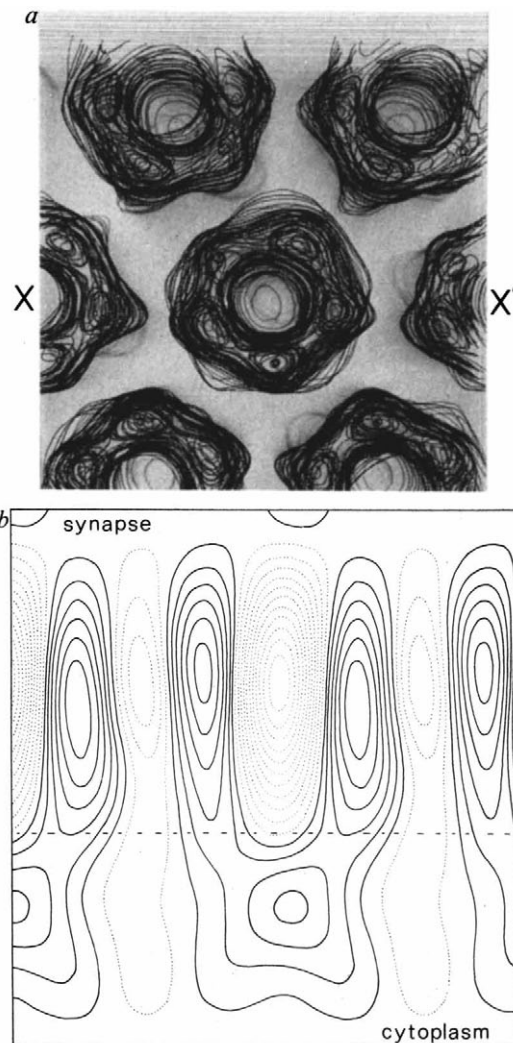


Fig. 4 a, Three-dimensional map of receptor molecules in the crystal lattice as they would appear viewed from the synaptic cleft; successive sheets are of sections parallel to the membrane plane separated by spaces corresponding to 5 Å. The centre-to-centre separation between receptors in the horizontal direction corresponds to 90 Å. b, Section along the line XX' in a, through the channel axes and perpendicular to the plane of the sheets; the central plane of the bilayer is indicated by the broken line. The phosphotungstate stain is excluded from the channel at a level ~ 15 Å above this line. Only contours corresponding to regions in which protein is concentrated are shown in a; all contours are shown in b. The outermost, full-line contours in a and b are close to the zero contour in Fig. 5 and represent the molecular boundary consistent with the volume calculated from protein composition.

synaptic and cytoplasmic portions it was 550 and 375 Å², respectively. As an approximate check on the value for the cross-section in the bilayer, a three-dimensional map was also calculated with approximate terms included for the central (0, 0) lattice line. These terms were derived from transforms of the striated border regions of tubes, as in Fig. 1a, using the (1, 2) lattice line (associated with the striation periodicity) to set the scale for the amplitudes and the origin of the phases. From the molecular boundary now on an absolute scale, a mean value of 500 Å² was obtained for the subunit cross-section in the bilayer, in reasonable agreement with the estimate based on protein composition.

Pentagonal symmetry

The five subunits are arranged regularly around the central axis of the receptor for much of its length, so that the structure has approximate pentagonal symmetry about this axis. The r.m.s. departures from exact symmetry are, respectively, 1.2 and 1.4 Å in radius and azimuth averaged over the portion of the protein

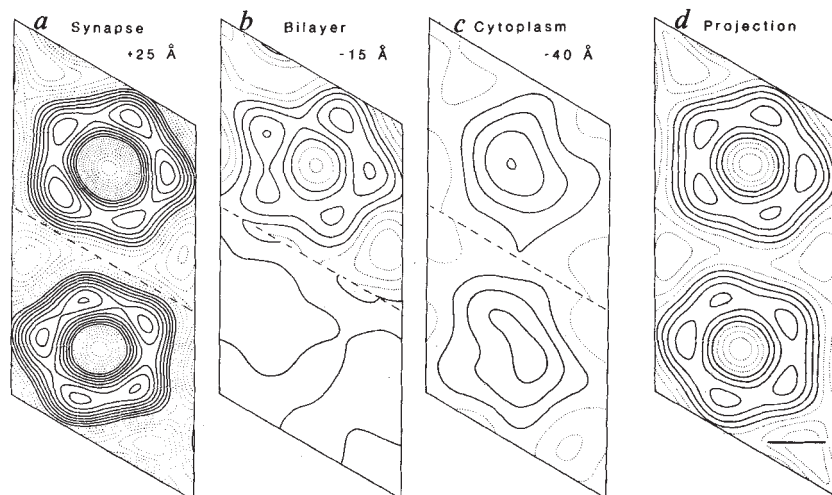


Fig. 5 *a-c*, Sections at different levels through a unit cell from the three-dimensional maps; *d*, the projected structure in frozen solution. The pair of receptors shown is thought to correspond to the δ - δ subunit linked physiological dimer¹³. In *a-c* the regions and levels indicated relate to Fig. 3: contours in the upper and lower halves refer to the frozen and the stained tubes, respectively; the two halves are related by symmetry as in *d*. The positive and zero contours are indicated by continuous lines, the negative contours by dotted lines; data from the (0, 0) lattice line, not included in the Fourier syntheses, have only a minor effect on the position of the zero contour. The scale bar (bottom right) corresponds to 25 Å.

in the hydrophobic core of the bilayer, and 0.8 and 1.9 Å averaged over the contiguous ~30-Å portion on the synaptic side. The total extent of this symmetrical region comprises about half of the length of the molecule, sufficient to dominate its appearance in projection (Fig. 5*d*). Nearer to the extremities, particularly on the cytoplasmic side, the pentagonal symmetry is not as marked, suggesting more profound structural differences between individual subunits in these regions? On the cytoplasmic side the distribution of mass about the axis of the oligomer is asymmetrical (Fig. 5*c*) and the cross-section is smaller, as if the parts of the subunits here have different lengths.

Discussion

The tubular crystals have led to a quantitative description of the quaternary structure of the acetylcholine receptor in a natural lipid-containing environment. The receptor is a cylindrical molecule having almost constant lateral dimensions (mean diameter ~65 Å) except for the part on the cytoplasmic side of the bilayer, which is narrower. It is composed of five rod-shaped subunits, each up to ~140 Å long. They lie approximately perpendicular to the membrane plane and are asymmetrically placed in relation to the bilayer, having two to three times more of their combined mass on the synaptic than on the cytoplasmic side. Subunits are similar in cross-section and are arranged symmetrically around a central opening, the presumed channel through which cations diffuse. The channel is partitioned equally into a wide (25-30-Å diameter) portion at the synaptic end of the structure and an unresolved narrow portion extending through the membrane into the interior of the cell.

In previous studies, electron images of stained molecules, viewed *en face* and from the side, have been used to deduce the overall shape of the protein and channel^{16,17,25-27}. Morphological observations have also been combined with X-ray or neutron diffraction data to obtain models for the cylindrically averaged distribution of mass^{18,19}. Seen most directly in frozen solution, the molecule appears more elongated and more angular in shape than has generally been supposed.

Identification of the membrane-spanning portion of the structure (Fig. 3) partitions the protein mass on either side in a way that is consistent with models inferred from the amino-acid sequences. However, the probable correctness of different models cannot be evaluated on this basis because of the possibility that the extrinsic cytoplasmic protein of apparent molecular mass 43,000 (ref. 20), thought to be present in the tubes, is bound regularly to the crystallized receptor and contributes some extraneous mass.

A striking aspect of the entire oligomer, previously suspected¹ but not seen until the present study, is that the subunits are arranged in almost equivalent positions around the channel, conferring on the structure an apparent high degree of pentagonal symmetry. This is most exact in the lipid bilayer and

over a contiguous ~30 Å region on the synaptic side, and may be even more extensive than observed because minor errors in our data would make it less pronounced. Such symmetry, combined with the observed amino-acid sequence homology, makes it probable that the individual subunits share common features of tertiary organization over the lengths that they are in contact. These properties could be important in facilitating a coordinated movement of all subunits around the channel, such as may occur in switching between different transitional states⁸.

The resolution of the analysis described here is insufficient to elucidate the internal structure of the subunits, but the observed details are consistent with models speculating that their membrane-spanning portions are composed of equal numbers of closely packed α -helical rods, or other configurations of the polypeptide chain²⁸. The value suggested for the subunit cross-section in this region, 500-540 Å², favours models with five α -helices extending through the membrane^{10,12} (assuming, for example, a packing similar to that in bacteriorhodopsin²¹).

We thank Dr P. Swann, Gatan Inc., for making available an electron microscope cold holder, and Dr C. Cazaux and his colleagues at the Institut de Biologie Marine, Arcachon for supplying *T. marmorata*. The research was supported by EMBO and Philippe Foundation Fellowships to A.B. and by a grant from the NIH.

Received 17 January; accepted 19 April 1985.

- Changeux, J.-P. *Harvey Lect.* **75**, 85-254 (1980).
- Karlin, A. in *The Cell Surface and Neuronal Function* (eds Cotman, C. W., Poste, G. & Nicolson, G. L.) 191-260 (North-Holland, Amsterdam, 1980).
- Conti-Tronconi, B. M. & Raftery, M. A. *Rev. Biochem.* **51**, 491-530 (1982).
- Noda, M. *et al. Nature* **299**, 793-797 (1982).
- Noda, N. *et al. Nature* **301**, 251-255 (1983).
- Claudio, T., Ballivet, M., Patrick, J. & Heinemann, S. *Proc. nat. Acad. Sci. U.S.A.* **80**, 1111-1115 (1983).
- Devillers-Thiery, A., Giraudat, J., Bentabollet, M. & Changeux, J.-P. *Proc. nat. Acad. Sci. U.S.A.* **80**, 2067-2071 (1983).
- Popot, J.-L. & Changeux, J.-P. *Physiol. Rev.* **64**, 1162-1239 (1984).
- Noda, M. *et al. Nature* **302**, 528-532 (1983).
- Finer-Moore, J. & Stroud, R. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 155-159 (1984).
- Lewis, C. A. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6110-6113 (1983).
- Guy, H. R. *Biophys. J.* **45**, 249-261 (1984).
- Brisson, A. & Unwin, P. N. T. *J. Cell Biol.* **99**, 1202-1211 (1984).
- Amos, L. A., Henderson, R. & Unwin, P. N. T. *Prog. Biophys. molec. Biol.* **39**, 183-231 (1982).
- Finch, J. T. & Klug, A. *J. molec. Biol.* **13**, 1-12 (1965).
- Cartaud, J., Benedetti, E. L., Sobel, A. & Changeux, J.-P. *J. Cell Sci.* **29**, 313-337 (1978).
- Sealock, R. *J. Cell Biol.* **92**, 514-522 (1982).
- Klymkowsky, M. W. & Stroud, R. M. *J. molec. Biol.* **128**, 319-334 (1979).
- Wise, D., Karlin, A. & Schoenborn, B. P. *Biophys. J.* **28**, 473-496 (1979).
- Froehner, S. C. *et al. Proc. nat. Acad. Sci. U.S.A.* **78**, 5230-5234 (1981).
- Henderson, R. & Unwin, P. N. T. *Nature* **257**, 28-32 (1975).
- Agard, D. A. *J. molec. Biol.* **167**, 849-852 (1983).
- Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A. & Homo, J.-C. *J. Microsc.* **128**, 219-237 (1982).
- Thon, F. Z. *Naturforsch.* **219**, 476-478 (1966).
- Kistler, J. K. & Stroud, R. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3678-3689 (1981).
- Zingsheim, H. P., Neugebauer, D.-C., Barrantes, F. J. & Frank, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 952-956 (1980).
- Bon, F. *et al. J. molec. Biol.* **176**, 205-237 (1984).
- Criado, M., Hochschwender, S., Sarin, V., Fox, J. L. & Lindstrom, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2004-2008 (1985).

Direct evidence using tritium data for throughflow from the Pacific into the Indian Ocean

Rana A. Fine

Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

Basin-wide exchange between the Pacific and Indian oceans through the Indonesian Archipelago has received attention, both in relation to the oceans' role in the Southern Oscillation¹ and in efforts to balance the salt and mass fluxes of the individual basins²⁻⁵. Wyrski⁶ made the first indirect estimate using hydrographic data of a mean annual transport from the Pacific to the Indian Ocean of $2 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ in the upper 200 m. Recent estimates have been considerably higher: $14 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ by Piola and Gordon², $10 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ by Godfrey and Golding³, and $5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ by Godfrey and Ridgeway³. However, Wunsch *et al.*⁴ concluded that, as a result of inverse calculations on sections across the South Pacific, there was a negligible throughflow transport. A meridional maximum in bomb-produced tritium (half-life 12.4 yr) observed in the South Equatorial Current (SEC) of the Indian Ocean is used here to show direct evidence that in the mean there is a net transport (in the upper 300 m) of $5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ from the Pacific into the Indian Ocean.

The use of bomb tritium, which is found in the oceans in the form of HTO, has become increasingly popular⁷⁻⁹ as a conservative tracer for oceanic circulation processes with decadal time scales. Decay-corrected tritium data collected during the Geochemical Ocean Sections (GEOSECS) Expeditions¹⁰ are anomalously higher in the South Indian than either the South Atlantic or South Pacific oceans. The presence of the Indian subcontinent at 25° N rules out a North Indian Ocean source, the asymmetric fallout pattern^{11,12} rules out a Southern Hemisphere source. Toggweiler and Trumbore¹³ found anomalously high bomb-produced ^{90}Sr in Indian Ocean corals which they attribute to throughflow of North Pacific water.

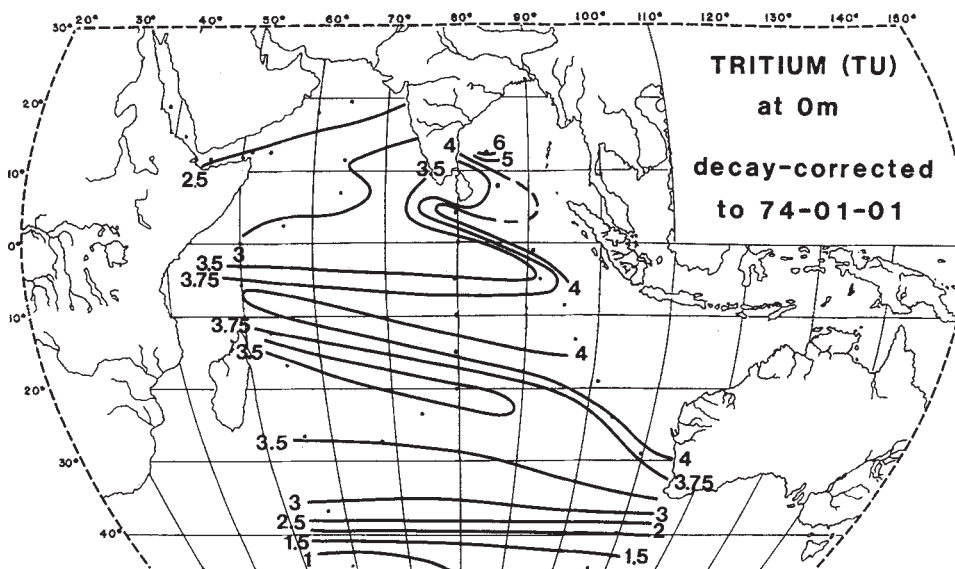
Figure 1 is a map showing tritium at the surface of the Indian Ocean¹⁰. Pre-bomb surface values may be considered to be $<1 \text{ TU}$ ($1 \text{ TU} = 10^{18} \times \text{mole fraction of } (^3\text{H}/^1\text{H}))$. Experimental error is $<0.1 \text{ TU}$. Tritium values are highest ($>6 \text{ TU}$) in the Bay of Bengal as result of runoff; yet the high values are confined to a shallow surface layer. Contrast the Arabian Sea where the highest evaporation in the Indian Ocean occurs, surface waters there contain 2.5 TU . In the Southern Hemisphere, the highest tritium values are found in the westward-flowing SEC located

between 10° and 20° S . Tritium $>3.5 \text{ TU}$ is found from about 5° N to 30° S , and it coincides fairly well with an intrusion of low salinity observed throughout the year¹⁴. The tritium contours in Fig. 1 generally follow an east-west pattern because the first order transport of water is advection by the zonal wind-driven currents.

Figure 2 is a map showing tritium at 200 m in the thermocline of the Indian Ocean¹⁰. Tritium values have decreased to $<1 \text{ TU}$ in the Bay of Bengal showing that high tritium in the runoff there is confined to a shallow surface layer. Its effect is only local because of dilution during subsequent mixing with low ambient levels of tritium. The zonal patterns at 200 m are similar to those at the surface; highest values are shifted slightly polewards. This is probably due to poleward contraction of the gyre with depth¹⁵, and perhaps also downwelling south of 15° S . Thus, at the time of sampling, bomb tritium had already mixed well into the thermocline of the subtropical gyre resulting in a maximum in the region of the SEC as a prominent feature down to 300 m. Most of the transport in the SEC occurs above 300 m (ref. 14).

The intrusion of low salinity/high tritium water into the South Indian Ocean is due to a combination of throughflow from the North and South Pacific, and precipitation⁶ which is very high near Indonesia. In general, twice as much tritium makes its way into the oceans through molecular exchange as through rain^{12,16}. In regions where precipitation and runoff are high, unusually low salinities and high tritium will be found, for example, the Bay of Bengal. However, the high tritium in the Bay of Bengal is confined to a shallow surface layer and has only a local effect. In addition, the tritium measured in precipitation in the few years before the GEOSECS Indian Ocean expedition was twice as high over the Bay of Bengal as that measured in precipitation over Indonesia¹⁷. This would result in even lower surface ocean tritium in the component from precipitation than in the Bay of Bengal runoff water. Thus, precipitation is not the main source for the low salinity/high tritium intrusion.

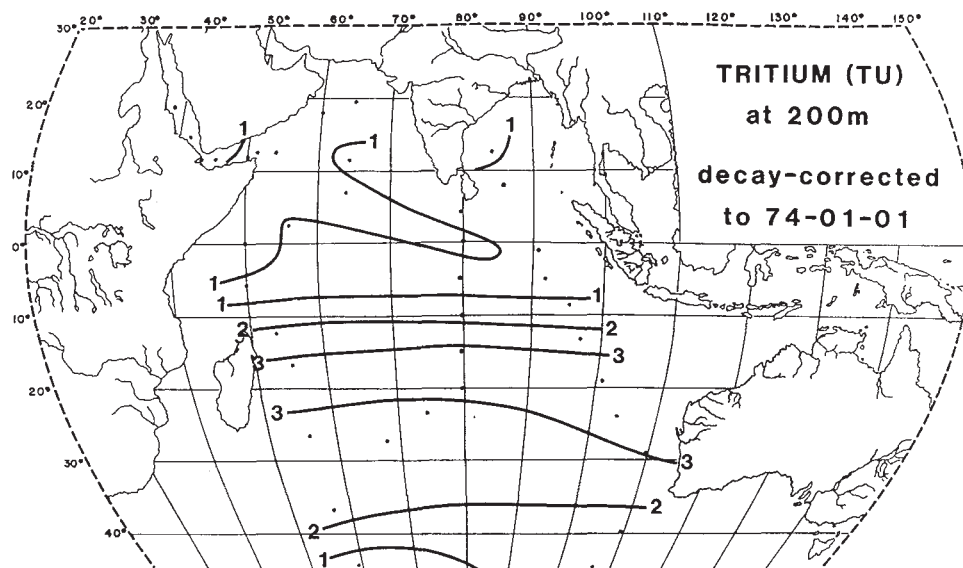
Wyrski⁶ concluded from hydrographic data that, even though the surface currents in the Indonesian Archipelago are subject to monsoonal variations, in all layers (except for the oxygen minimum) movement of water from the Pacific (via the Timor Current) to the Indian predominates. The transport has decreasing intensity with depth. Although the upper waters in the western tropical South Pacific contribute to the throughflow, they are too low in tritium⁸ and high in salinity¹⁸ to be a major source. On the other hand, GEOSECS tritium in the Pacific North Equatorial Current (NEC)⁸ was twice as high as the South Indian Ocean. Maps by Wyrski⁶ and Tsuchiya¹⁴ show some of the westward-flowing NEC turning south near Mindano, both feeding into the North Equatorial Counter Current and



© 1985 Nature Publishing Group

Fig. 1 Map of tritium (TU) at the surface of the Indian Ocean decay-corrected to 74-01-01. Data were collected as part of the GEOSECS program and analysed by Ostlund and Brescher¹⁰. •, Station locations.

Fig. 2 Map of tritium (TU) at 200 m in the Indian Ocean decay-corrected to 74-01-01. Data were collected as part of the GEOSECS program and analysed by Ostlund and Brescher¹⁰ (1982). •, Station locations.



entering the Indonesian Archipelago. Temperature and salinity near Mindano correlate with those in the western basins of the archipelago^{5,6}, in agreement with the idea of the North Pacific being a major source for the low salinity/high tritium intrusion in the Indian Ocean.

The additional information obtained when tritium is used as a tracer for oceanographic processes is particularly important for this problem because it provides an integrated picture of the decadal mean circulation. Tritium data¹⁰ are combined with annual averages of temperature and salinity¹⁹ into a box model to calculate the net transport from the Pacific into the Indian Ocean. The meridional extent of the box is chosen from 10 to 20° S and the depth range is 0–300 m; that region is fairly homogeneous in water mass properties and it encompasses most of the transport of the SEC. The eastern and western extent of the box are determined by geography—the Indonesian Archipelago and Madagascar.

The box model calculation²⁰ assumes that mass and chemical balances within the box are entirely consistent with pure advection. The first-order physics are:

$$u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} = \frac{\partial C}{\partial t} - \lambda(C) \quad (1)$$

where according to convention u is the advection in the east (x)/west direction, v is the advection in the north (y)/south direction, C is the concentration, t is time, and λ is the decay term for tritium. If equation (1) is integrated over the volume (V) of the box then:

$$C_N v_N A_N + C_S v_S A_S + C_E u_E A_E = \frac{\Delta C}{\Delta t} V - \lambda C V \quad (2)$$

where the C_i s represent the average concentrations and the A_i s represent the areas along the boundaries of the box. Table 1 gives the actual values for C_i s which were obtained by integrating the data^{10,19} observed along the boundaries of the box. Equation (2) is written for volume, salinity, temperature and tritium, giving a total of four equations and three unknowns. Using a least squares regression it may be solved for the advective velocities: v_N , v_S and u_E .

The air-sea exchange ($\Delta C/\Delta t$) may be evaluated using literature values. The net freshwater loss to the atmosphere by the Indian Ocean between 10 and 20° S calculated by Baumgartner and Reichel²¹ is $0.14 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. The heat flux between 10 and 20° S in the Indian Ocean estimated by Talley²² was a net loss from the ocean to the atmosphere of 2 W m^{-2} .

Rates of addition of tritium by precipitation and vapour exchange to the South Indian Ocean are given in 5° latitude bands by Weiss and Roether¹². When averaged over the time of

the bomb transient, the deposition rate for the region 10–20° S is balanced by tritium decay to within the 30% error given for the deposition rate. Although subsurface observations of tritium are only available from the GEOSECS expedition for the Indian Ocean, there are observations spanning at least 10 yr in the NEC, which is the source region for the low salinity/high tritium intrusion. After 1974, decay has balanced input in the NEC²³, because we were already on the flat part of the deposition curve. Therefore, it is reasonable to consider the surface and subsurface tritium in a steady state.

The results of the regression together with the statistical errors in the coefficients are given in Table 1. The net transport into the box across the eastern boundary corresponds to the through-flow from the Pacific into the Indian Ocean; it is $5.1 \pm 0.4 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. There is a small net northwards transport across 10° S (out of the box) of $-2.4 \pm 0.1 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ and a small net southwards transport across 20° S (out of the box) of $-2.6 \pm 0.4 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. The north/south transports are difficult to assess physically as they encompass the surface Ekman poleward flow and part of the equatorward convergent geostrophic flow, as well as the branching of the SEC at its western terminus. A possible explanation for the integrated net northwards flow across 10° S could be related to the fact that Madagascar extends only to 10° S. Schott (personal communication) has recently found a strong northwestwards current at the northern tip of Madagascar.

Using the same data as described above, the model was run for a box 0–100 m deep encompassing the mixed layer. The

Table 1 Average concentrations and resultant net transports calculated from a box model for the region 10–20° S in the Indian Ocean

	N(10° S)	S(20° S)	E
0–300 m			
C (temperature, °C)	18.57	20.61	20.42
C (salinity, ‰)	34.85	35.39	34.43
C (tritium, TU)	2.4	2.5	3.0
Area (10^8 m^2)	20.7	19.8	2.9
Transport ($10^6 \text{ m}^3 \text{ s}^{-1}$)	-2.4	-2.6	+5.1
Error ($10^6 \text{ m}^3 \text{ s}^{-1}$)	±0.1	±0.4	±0.4
0–100 m			
C (temperature, °C)	25.66	24.30	25.70
C (salinity, ‰)	34.80	35.20	34.20
C (tritium, TU)	3.5	3.6	3.8
Area (10^8 m^2)	6.9	6.6	1.0
Transport ($10^6 \text{ m}^3 \text{ s}^{-1}$)	±1.7	-3.6	+2.0
Error ($10^6 \text{ m}^3 \text{ s}^{-1}$)	±0.7	±0.2	±0.5

Positive (negative) transports indicate a net flow into (out of) the box across that boundary.

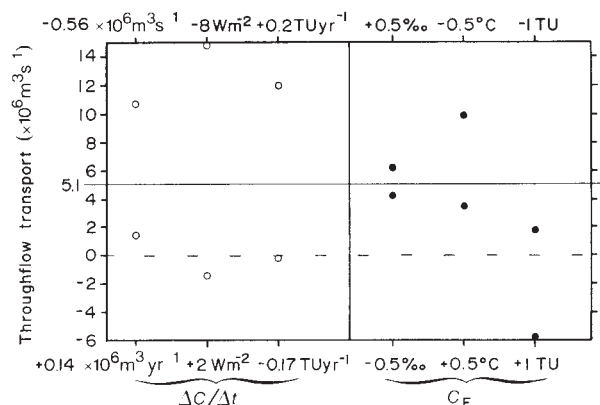


Fig. 3 Comparison of the box model result of $5.1 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ throughflow to variations in: ○, air-sea flux parameters ($\Delta C/\Delta t$) ($\Delta C/\Delta t$ is varied for: freshwater, heat, and tritium input); and ●, concentrations along the eastern boundary of the box (C_E) (C_E is varied for salinity, temperature, and tritium).

results of that regression are also included in Table 1. The net transports across both 10° S ($1.7 \pm 0.7 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ into the box) and 20° S ($-3.6 \pm 0.2 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ out of the box) are southwards; this is consistent with theories of a surface Ekman poleward flow. There is a net transport into the box across the eastern boundary of $2.0 \pm 0.5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. This calculation, which gives only $2 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ for the throughflow (0–100 m), shows that more than half of the transport calculated for the box 0–300 m is below the mixed layer.

Figure 3 shows the sensitivity of the throughflow transport to uncertainties which are at least several times the errors in the data. At either extreme in the balance between the tritium decay rate (-0.17 TU yr^{-1}) and deposition rate¹² (0.20 TU yr^{-1}), the effect is to decrease to nearly zero or double the throughflow transport. The uncertainties in the freshwater and heat air-sea fluxes are unknown, although it is unlikely that the directions of the fluxes are unknown. A fourfold increase in these fluxes results in more than a doubling of the throughflow transport. Varying concentrations along the eastern (Fig. 3), northern and southern boundaries of the box result in net positive throughflows, except when C_E and C_N for tritium are increased by 30% (1 TU). These sensitivity tests always result in a net northward transport across 10° S ; this is consistent with observations of a northward current at the northern tip of Madagascar (F. Schott, personal communication). The tests also result in a positive correlation between throughflow and southwards transport across 20° S ; this is consistent with the assertion by Piola and Gordon² that the upper layer throughflow transport must be balanced by southwards flow across 35° S in the Indian Ocean.

A throughflow transport exceeding $10 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ is unlikely. Godfrey and Golding⁵ pointed out this high a transport would require large upper-layer velocities $> 1 \text{ m s}^{-1}$ in some of the Indonesian passages; these have not been observed. In addition, the throughflow would then amount to 30–50% of the total volume transport for the SEC which is $30\text{--}40 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ (F. Schott, personal communication). Because of the high tritium intrusion, it is also unlikely that there is no net throughflow transport or even that it is reversed. When the regression was run for three equations and three unknowns, successive omission of salinity, heat and tritium gave throughflows respectively of 2.0 , 13.4 and $10.9 \times 10^6 \text{ m}^3 \text{ s}^{-1}$; direction of the transports across the other two boundaries remained unchanged. Salinity has the most redundancy, that is freshwater, heat and tritium are the three most linearly independent parameters. These sensitivity analyses suggest that the box model calculation of $5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ for the throughflow transport is probably good to within $\pm 50\%$.

The tritium data show a direct link between the Pacific and Indian oceans. The interannual variability in the throughflow could have direct implications for climate because it is coincident with a warm pool of 28° C or higher¹⁴. The warm pool is found to extend from the western Pacific through to the Indian

Ocean. The unique relationship of the Southern Oscillation to sea-surface temperature (SST) in the region extending from the warm pool southwards to northern Australia^{24,25} has been shown to have a high correlation²⁶. In addition, SSTs in that region have been shown to have an influence on the monsoon^{27,28}, because the largest region of deep atmospheric convection occurs over the warm pool. Changes in intensity and location of this convection zone are related to variability in both the global atmospheric circulation and the monsoon, and are probably also influenced by changes in the underlying warm, low-salinity throughflow water.

I thank Gary Meyers for suggesting the examination of the tracer data in relation to the throughflow problem; H. Gote Ostlund for the tritium data; Fritz Schott for suggesting the use of the air-sea flux terms; James Mattson, Julian McCreary, Claes Rooth for many useful comments; and the NSF OCE7822120 and OCE8024904 for support.

Received 4 December 1984; accepted 2 April 1985.

- Bye, J. A. T. & Gordon, A. H. *Nature* **296**, 52–54 (1982).
- Piola, A. R. & Gordon, A. L. *J. phys. Oceanogr.* **14**, 747–753 (1984).
- Godfrey, J. S. & Ridgeway, K. R. *Trop. Ocean-Atmos. Newslett.* **26**, 16–17 (1984).
- Wunsch, C., Hu, D. & Grant, B. *J. phys. Oceanogr.* **13**, 725–753 (1983).
- Godfrey, J. S. & Golding, T. J. *J. phys. Oceanogr.* **11**, 771–779 (1981).
- Wyrtki, K. *Physical Oceanography of the Southeast Asian Waters*, NAGA Rep. No. 2 (Scripps Institution of Oceanography, Washington DC 1961).
- Jenkins, W. J. *J. mar. Res. Suppl.* **40**, 265–290 (1982).
- Fine, R. A., Reid, J. L. & Ostlund, H. G. *J. phys. Oceanogr.* **11**, 3–14 (1981).
- Sarmiento, J. L., Rooth, C. G. H. & Roether, W. *J. geophys. Res.* **87**, 8047–8056 (1982).
- Ostlund, H. G. & Brescher, R. *GEOSECS Tritium, Tritium Laboratory Data Rep. No. 12* (University of Miami, Rosenstiel School of Marine and Atmospheric Science, 1982).
- Fine, R. A. & Ostlund, H. G. *Geophys. Res. Lett.* **4**, 461–464 (1977).
- Weiss, W. & Roether, W. *Earth planet. Sci. Lett.* **49**, 435–446 (1980).
- Toggweiler, J. R. & Trumbore, S. *Earth planet. Sci. Lett.* (submitted).
- Wyrtki, K. *Oceanography Atlas of the International Indian Ocean Expedition* (NSF, Washington DC, 1971).
- Reid, J. L. *Evolution of Physical Oceanography* (eds Warren, B. A. & Wunsch, C.) 70–111 (MIT Press, 1981).
- Craig, H. & Gordon, L. I. *Conf. Stable Isotopes in Oceanogr. Studies and Paleotemperatures*, Pisa (Consiglio Nazionale Delle Ricerche, Laboratorio Di Geologia Nucleare, 1965).
- Statistical Treatment of Environmental Isotope Data in Precipitation*, Tech. Rept. No. 206 (IAEA, Vienna, 1981).
- Tsuchiya, M. *Johns Hopkins Oceanogr. Stud.* **4**, (1968).
- Levitus, S. *Climatological Atlas of the World Ocean* (NOAA, Rockville, 1982).
- Wunsch, C. *J. geophys. Res.* **89**, 7971–7978 (1984).
- Baumgartner, A. & Reichel, E. *The World Water Balance—Mean Annual Global, Continental and Maritime Precipitation, Evaporation and Runoff* (Elsevier, Amsterdam, 1975).
- Talley, L. D. *J. phys. Oceanogr.* **14**, 231–241 (1984).
- Fine, R. A., Peterson, W. H. & Ostlund, H. G. *J. phys. Oceanogr.* (submitted).
- Koninklijk Nederlands Meteorologisch Instituut. *Indische Oceaan. Oceanografische en Meteorologische Gegevens* 2nd edn, Publ. No. 135 (De Bilt, 1952).
- Nicholls, N. *Trop. Ocean-Atmos. Newslett.* **24**, 11–12 (1984).
- Wright, P. *Trop. Ocean-Atmos. Newslett.* **20**, 15–16 (1983).
- Gadgil, S., Joseph, P. V. & Joshi, N. *J. Trop. Ocean-Atmos. Newslett.* **27**, 7–8 (1984).
- Nicholls, N. *Nature* **306**, 576–577 (1983).

Importance of thrust faulting in the tectonic development of northern Victoria Land, Antarctica

G. M. Gibson

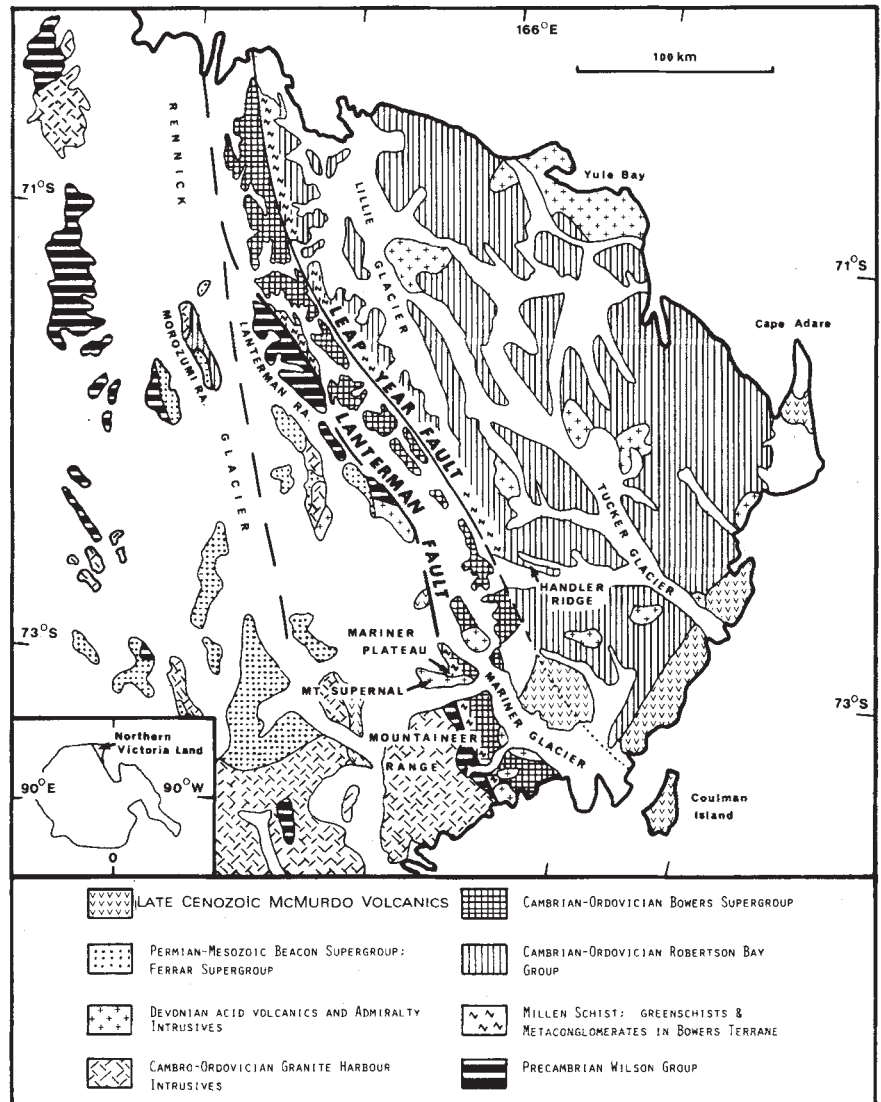
School of Applied Science, Darling Downs Institute, Toowoomba, Queensland 4350, Australia

T. O. Wright

Earth Science Division, National Science Foundation, Washington, DC 20550, USA

The present configuration of Wilson Group, Bowers Supergroup and Robertson Bay Group in northern Victoria Land (Fig. 1) has been attributed to high-angle block-faulting^{1–3} or more recently⁴ to major strike-slip faulting along the Lanterman and Leap Year Fault zones. In this latter interpretation⁴, adjacent units have allochthonous relationships and constitute separate geological terranes. We offer additional support for the concept of allochthonous terranes in northern Victoria Land but suggest here an alternative

Fig. 1 Geological map of northern Victoria Land (compiled from data in refs 1, 3, 4, 12).



model to explain terrane juxtapositions. Our model emphasizes the role of thrust tectonics and implies severe structural telescoping of entire terranes with significant loss of intervening crust. We further suggest that this occurred during the Lower Palaeozoic Ross Orogeny in response to the collision of Antarctica with a westwardly directed eastern continental block. This requires that the northern Victoria Land segment of Gondwana was a zone of major plate convergence throughout the Cambrian-early Ordovician period.

The Wilson terrane contains mainly (?Precambrian) amphibolite facies gneisses intruded by Lower Palaeozoic S-type granitoids (Granite Harbour Intrusives^{5,6}). These gneisses have undergone four phases of deformation⁷⁻⁹, the first two of which were accompanied by intense flattening and development of boudinage. A conspicuous stretching lineation defined by elongated mineral aggregates is widespread throughout the Lanterman Range and also occurs in Wilson Group gneisses immediately west of the Bowers Supergroup in the Mountaineer Range. In the Lanterman Range, this lineation originated during the second phase of deformation and everywhere plunges steeply down-dip within the north-west-trending regional foliation (Fig. 2a). The third and fourth phases of deformation in the Lanterman Range were associated with retrogressive metamorphism to the greenschist facies and the widespread development of crenulation structures^{7,8}.

The Bowers terrane contains a Cambrian-early Ordovician regressive sequence^{2,10} which has at its base the Sledgers Group. Sledgers Group comprises two mutually interdigitating units¹⁰: the Molar Formation composed of marine volcanogenic sedi-

ments and the Glasgow Volcanics consisting of pillow lavas, basaltic and andesitic breccias, and subordinate dacite-rhyolite^{4,10}. Basic rocks within the Glasgow Volcanics are interpreted⁴ as primitive island arc tholeiites.

Overlying^{2,10} the Sledgers Group are the non-volcanogenic, shallow or marginal marine, Mariner Group and the non-marine Leap Year Group.

Two other units are also included here in the Bowers terrane; the Lanterman and Husky Conglomerates, both of which occur only in the Lanterman Range along the western margin of the terrane. They are characterized⁷ by intensely flattened pebbles and provide important clues about the nature of tectonism along the contact between the Bowers and Wilson terranes. Other conglomerates with severely flattened, or sometimes even stretched, pebbles occur in the Molar Formation immediately south of the Meander Glacier in the Mountaineer Range^{3,11,12}.

The Bowers Supergroup was mainly metamorphosed under prehnite-pumpellyite or pumpellyite-actinolite facies conditions¹³ and generally underwent folding about gently plunging, north-west-trending axes^{13,14}. Multiply-deformed greenschist facies rocks are restricted^{7,11,12} to a narrow (2-4 km) belt immediately east of the contact with Wilson Group (Fig. 1). Highly strained conglomerates in this belt exhibit characteristics of both flattening and constrictional deformation^{7,11}. Fold axes within this belt also show extremely variable plunge, sometimes steepening quite markedly over short distances, although they always maintain a constant north-west orientation within their axial surfaces (Figs 2b, d-f). This suggests non-cylindrical folding, possibly even the presence of sheath folds. Where

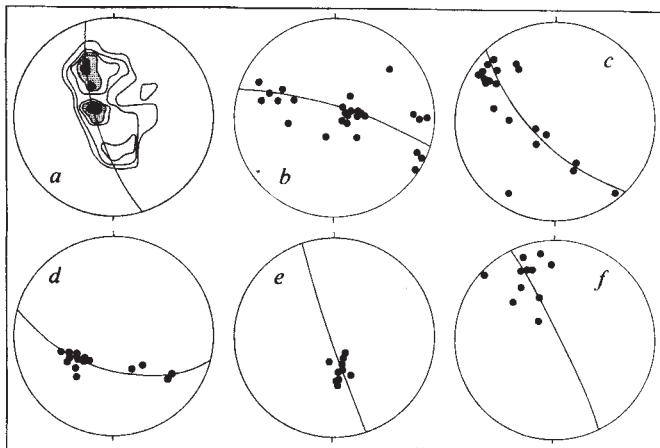


Fig. 2 Equal area stereographic projections of regional structures in the different terranes of northern Victoria Land. Solid lines, dominant schistosity or cleavage; dots, lineations and fold axes except in *a* which includes 112 stretching lineations from Wilson Group gneisses in the Lanterman Range. Contour interval is 1%, 3%, 6%, 9% and 12% per 1% area. *b*, Bowers terrane immediately east of Wilson terrane in the Lanterman Range. *c*, Robertson Bay terrane, data from ref. 14. *d-f*, Bowers terrane; *d*, belt of greenschist facies rocks immediately east of contact with Wilson terrane; *e*, greenschist facies rocks on the Mariner Plateau; *f*, virtually unmetamorphosed Bowers supergroup from area in the lower Mariner glacier. Shading accentuates contour interval.

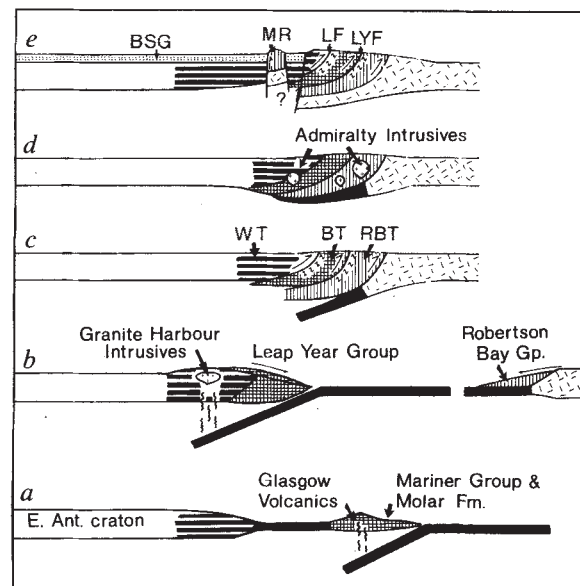


Fig. 3 Tectonic model for northern Victoria Land. *a*, Middle-late Cambrian: development of Glasgow Volcanic arc concomitant with the deposition of Molar Formation and the Mariner Group in a forearc basin. To the east, Robertson Bay Group is deposited as a clastic wedge with continental provenance. *b*, Late Cambrian-early Ordovician: following accretion of the Glasgow Volcanic arc onto the Gondwana continental margin, the centre of magmatic activity migrated westward leading to the intrusion of the Granite Harbour Intrusives and the penecontemporaneous deposition of Leap Year Group (following uplift and erosion of granites). *c*, Ordovician: Ross Orogeny with substantial crustal thickening and underthrusting of one terrane beneath the next. Deformation and regional metamorphism with the Bowers Supergroup; retrogressive metamorphism within the Wilson terrane. *d*, Devonian-Carboniferous: intrusion of Admiralty Intrusives. *e*, Post-Ferrar Supergroup Rennick block faulting¹. Differential uplift and erosion in the Morozumi Range remove the overlying Wilson and Bowers Supergroup rocks to create a tectonic window in which Robertson Bay Group sediments are exposed. Abbreviations: BSG, Beacon Supergroup; BT, Bowers Terrane; LF, Lanterman Fault; LYF, Leap Year Fault; MR, Morozumi Range; RBT, Robertson Bay Terrane; WT, Wilson Terrane.

asymmetrical folds have been observed^{9,11}, they generally indicate tectonic transport to the north-east, at right angles to the structural trend of the Bowers terrane. Since the Leap Year Group contains only one-fold generation, and locally^{10,11} rests unconformably upon Sledgers Group (without intervening Mariner Group), some deformation of Bowers rocks evidently occurred prior to deposition of the Leap Year Group.

The Robertson Bay terrane contains mainly marine sediments of continental provenance^{1,2,15,16}. Along with Bowers Supergroup, these rocks were intruded by the Devonian-early Carboniferous Admiralty Intrusives which possess I-type characteristics^{3,17,18}. Although formerly considered to be Precambrian or possibly Cambrian in age, the recent discovery^{19,20} of Ordovician fossils in the Robertson Bay Group at Handler Ridge makes a Cambrian-early Ordovician age more likely.

The structural relations of the Robertson Bay Group and Bowers Supergroup are strikingly similar. To the east, weakly metamorphosed^{21,22} Robertson Bay Group contains^{14,21,23} a single phase of upright, variably plunging folds which trend north-west (Fig. 2*c*), whereas close to the Leap Year Fault there is a belt of multiply deformed greenschist facies rocks (Millen schists)^{23,24}. These schists occupy a structural position comparable with that of the greenschist facies rocks along the western margin of the Bowers terrane. Both belts lie immediately east of major tectonic boundaries: the Leap Year and Lanterman Faults respectively. This implies a genetic as well as spatial relationship between the faults and greenschist facies rocks.

The Leap Year Fault has been traced^{3,25,26} across virtually the entire length of northern Victoria Land. Recent mapping¹¹ suggests the Lanterman Fault (type locality^{14,25} in the Lanterman Range) extends southwards into the Mountaineer Range. In both areas, Wilson Group rocks are thrust^{7,11} against the Bowers terrane (Fig. 1). Moreover, on the Mariner Plateau (Fig. 1), the Lanterman Fault is cut by one of the Devonian Admiralty Intrusives (Mt Supernal Granite). Mylonites in this same region partly recrystallized during the associated contact metamorphism. Because the country rocks had also undergone regional metamorphism and deformation before granite emplacement¹¹, we infer that both the thrust faulting and the regional deformation and metamorphism are manifestations of a pre-Devonian event: the Lower Palaeozoic Ross Orogeny.

The Lanterman and Leap Year faults both possess features compatible with major strike-slip faults. They have straight traces, dip steeply and are marked by wide crush zones, up to 500 m wide in the case of the Leap Year Fault³. These few features aside, there is no additional evidence for strike-slip faulting. Both faults are better interpreted as major thrusts. Along the entire length of the Lanterman Fault in both the Lanterman and Mountaineer ranges, Wilson Group amphibolite facies gneisses are consistently juxtaposed against lower grade Bowers Supergroup rocks. The faulted contact between these two units, clearly visible north of Mitten Spur in the Lanterman Range, dips steeply south-west. Slickensides along this fault indicate that movement was directed towards the north-east⁷. Parallel, and presumably related, thrust faults occur^{9,12} within the Mountaineer Metamorphics west of the Lanterman Fault. They too bring higher-grade rocks over lower-grade ones. A syn-metamorphic thrust fault has also been reported^{23,24} in the Millen Range just east of the Leap Year Fault.

In the western Lanterman Range, Lower Ordovician Granite Harbour Intrusives are thrust over Permian tillites and slate⁸. Late-stage reverse faults also affect the deformed conglomerates in the eastern Lanterman Range⁷. Some thrusting clearly post-dates the Ross Orogeny and may represent the reactivation of pre-existing structural discontinuities. Reorientation and/or modification of the earlier thrust faults seems likely.

Based on our interpretation that the Lanterman and Leap Year faults are major thrust boundaries, we present an alternative tectonic model for northern Victoria Land (Fig. 3). It is at variance with earlier ideas³ of a Bowers graben and precludes significant strike-slip faulting⁴ along the Lanterman and Leap Year faults during the early Palaeozoic. In our model, the present disposition of rock types is primarily the result of massive telescoping along the Lanterman and Leap Year faults with one terrane underplated beneath the next. Elements of a primitive island arc (Sledgers Group) were accreted onto the Gondwana margin before a continent-continent collision with significant loss of oceanic crust. The Robertson Bay Group represents an accretionary wedge brought into juxtaposition with the Bowers Supergroup from the east following closure of the intervening ocean. Our model not only implies allochthonous relationships between adjacent terranes but substantial crustal thickening beneath the Transantarctic Mountains during the Ross Orogeny.

Other attributes of our model include:

- (1) It provides different source areas for the compositionally dissimilar Sledgers and Robertson Bay Group sediments.
- (2) Structural telescoping and significant underthrusting of Bowers Supergroup below the Wilson terrane brought island-arc tholeiites (Glasgow Volcanics) within 6 km of the calc-alkaline Granite Harbour Intrusives and removes earlier objections⁴ that the Bowers terrane is too narrow for a typical arc.
- (3) Fold vergence, the common structural trend of all three terranes, and the down-dip stretching lineation in Wilson Group gneisses are consistent with thrusting directed towards the north-east.
- (4) Thrusting immediately following regional metamorphism and/or cessation of subduction emplaced hot Wilson Group rocks over the Bowers terrane. This produced higher temperatures (greenschist facies metamorphism) and the intense flattening observed in rocks along the western margin of the Bowers terrane. Similarly, the Millen schists east of the Leap Year Fault resulted from the thrusting of Bowers rocks over the Robertson Bay terrane.
- (5) The Robertson Bay and Bowers terranes are underlain by oceanic crust which provided possible source rocks for the Admiralty I-type granites.

A weakness in our model is the lack of an appropriate suture zone between the Bowers and Robertson Bay terranes. The Leap Year Fault is one possible choice but no fragment of a subduction complex, mélangé, blueschist terrane or ophiolite belt is preserved along its trace. An additional problem is the absence of continental crust east of the Robertson Bay terrane, although rocks of suitable age have recently been reported²⁷ on Surgeon Island in Yule Bay.

We thank the Bundesanstalt für Geowissenschaften und Rohstoffe for the opportunity to join GANOVEX III and for the invitation to participate in the 1984 NVL workshop in Hannover where many of these ideas were discussed.

Received 7 January; accepted 9 April 1985.

1. Tessensohn, F. *et al.* *Geol. Jb.* **B41**, 31–88 (1981).
2. Laird, M. G., Bradshaw, J. D. & Wodzicki, A. in *Antarctic Geoscience*, 535–542 (University of Wisconsin Press, Madison, 1982).
3. Stump, E. *et al.* *Nature* **304**, 334–336 (1983).
4. Weaver, S. D., Bradshaw, J. D. & Laird, M. G. *Earth planet. Sci. Lett.* **68**, 128–140 (1984).
5. Gunn, B. M. & Warren, G. *Bull. N.Z. geol. Surv.* **71**, 157 (1962).
6. Sturm, A. G. & Carryer, S. N.Z. *J. Geol. Geophys.* **13**, 408–435 (1970).
7. Gibson, G. M. *Geol. Jb.* **B60**, 117–141 (1984).
8. Roland, N. W., Gibson, G. M., Kleinschmidt, G. & Schubert, W. *Geol. Jb.* **B60**, 319–361 (1984).
9. Kleinschmidt, G., Roland, N. W. & Schubert, W. *Geol. Jb.* **B60**, 213–251 (1984).
10. Laird, M. G. & Bradshaw, J. D. in *Antarctic Earth Science*, 123–126 (Australian Academy of Science, Canberra, 1983).
11. Gibson, G. M., Tessensohn, F. & Crawford, A. R. *Geol. Jb.* **B60**, 289–318 (1984).
12. Tessensohn, F. *Geol. Jb.* **B60**, 371–396 (1984).
13. Wodzicki, A., Bradshaw, J. D. & Laird, M. G. in *Antarctic Geoscience*, 594–554 (University of Wisconsin Press, Madison, 1982).
14. Bradshaw, J. D., Laird, M. G. & Wodzicki, A. in *Antarctic Geoscience*, 809–816 (University of Wisconsin Press, Madison, 1982).
15. Field, B. D. & Findlay, R. H. in *Antarctic Earth Science*, 102–106 (Australian Academy of Science, Canberra, 1983).
16. Wright, T. O. *Geol. Jb.* **B41**, 127–138 (1981).
17. Wyborn, D. *Geol. Jb.* **B41**, 229–249 (1981).
18. Vetter, U. *et al.* in *Antarctic Earth Science*, 140–143 (Australian Academy of Science, Canberra, 1983).

19. Burrett, C. F. & Findlay, R. H. *Nature* **307**, 723–725 (1984).
20. Wright, T. O., Ross, R. J. Jr & Repetski, J. E. *Geology* **12**, 301–305 (1984).
21. Kleinschmidt, G. & Skinner, D. M. B. *Geol. Jb.* **B41**, 155–199 (1981).
22. Kleinschmidt, G. in *Antarctic Earth Science*, 119–122 (Australian Academy of Science, Canberra, 1983).
23. Findlay, R. H. & Field, B. D. in *Antarctic Earth Science*, 107–112 (Australian Academy of Science, Canberra, 1983).
24. Wright, T. O. & Findlay, R. H. *Geol. Jb.* **B60**, 105–116 (1984).
25. Dow, J. A. S. & Neall, V. E. *N.Z. J. Geol. Geophys.* **17**, 659–714 (1974).
26. Jordan, H. *et al.* *Geol. Jb.* **B60**, 57–81 (1984).
27. Vetter, U., Lenz, H., Kreuzer, H. & Besang, C. *Geol. Jb.* **B60**, 363–369 (1984).

Diagenesis of sediments beneath the Ross Ice Shelf and their sedimentary history

R. Raiswell

Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

M. Md. Tan

School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Micropalaeontological studies of sediment cores collected from beneath the Ross Ice Shelf, Antarctica, give conflicting stratigraphical interpretations. The diatom flora initially indicated an *in situ* succession of later middle Miocene age^{1,2}, which was supported by the occurrence of benthic foraminifera³. However, subsequent diatom analyses record the presence of flora with mixed ages (mostly Miocene but also Pliocene and Pleistocene) throughout both of the two lithological units present in the cores. These flora are taken to indicate repeated reworking, which most recently occurred no earlier than the late Pleistocene⁴. These contrasting views have since been re-affirmed^{5,6}. Our studies of sediment geochemistry show self-consistent patterns in organic carbon, organic nitrogen and sulphur diagenesis which suggest a Recent, *in situ* origin, at least for the upper unit.

The 47 Ross Ice Shelf Project (RISP) cores collected at site J9 (82°22' S, 168°28' W) in the 1977–78 field season have lengths ranging from 15 to 122 cm and contain an upper light olive-grey unit (9–30 cm thick) at the base of which is a thin (<1 cm) brown-orange, iron-rich layer, providing a sharp colour contrast with the darker olive-grey lower unit (>86 cm) thick³. Both upper and lower units consist of pebble-granule diatomaceous mud, soft and sloppy in the upper unit but firm and compact in the lower. Quartz, feldspars, clays and diatoms generally comprise >80% of the sediments and the only discernible mineralogical difference between the two units is the presence of minor chlorite and vermiculite in the upper unit⁷.

Siliceous diatoms are the most abundant microflora and occur uniformly throughout the upper and lower units. Calcareous benthic foraminifera also occur in the lower unit but are absent from the upper unit^{1–4}. There are two conflicting views as to the age of the microflora. One view is that the diatoms are well-preserved, except in the upper unit, and indicate a middle Miocene age with the foraminifera in the lower unit being slightly older (early to middle Miocene). This view is consistent with the presence of an *in situ* middle Miocene succession, with the upper lithological unit representing an alteration product of the underlying sediment^{2,7}. The alternative view crucially depends on the recognition of minor proportions of smaller, more delicate and poorly preserved Pliocene to late Pleistocene diatoms, which have been recorded amongst the predominantly middle Miocene flora throughout both units. The presence of flora with mixed ages is taken to indicate subglacial reworking, probably in the late Pleistocene (21,000–17,000 yr BP) and possibly also earlier⁶. Thus, there are controversial questions concerning the timing of deposition and the extent of reworking and alteration in the RISP cores, which are of importance in establishing the history of the West Antarctica Ice Sheet.

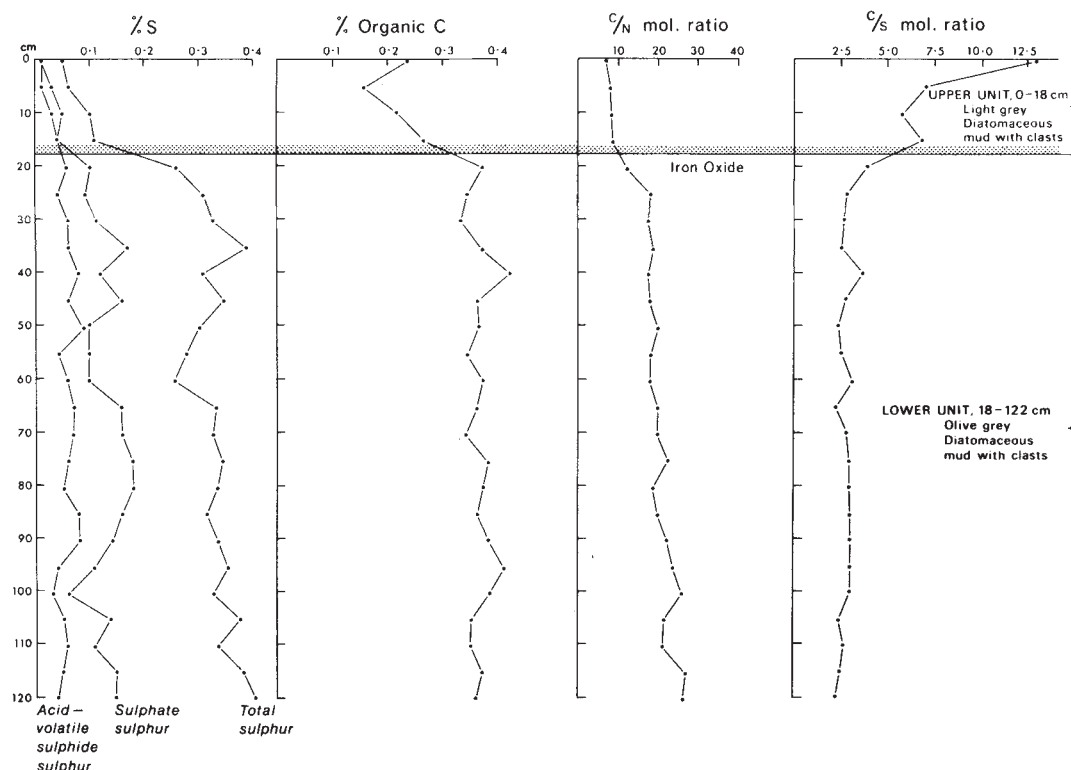


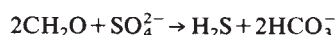
Fig. 1 Depth trends in S species, organic C and the C/S and C/N mole ratios in RISP core no. 12.

A study of chemical diagenesis was undertaken to clarify aspects of the sedimentary history. The longest RISP core, no. 12, was selected to define chemical gradients most accurately. Excess water was allowed to drain away before cores were frozen for transportation³. However, samples were received in a dry, unfrozen state and chemical effects may arise from sulphide oxidation and contamination by salts from dried porewater: the latter effect is small. This core provided 25 samples located at 5-cm intervals over 122 cm. These were crushed to <90 µm and analysed for organic C, organic N (by Hewlett Packard CHN analyser), acid-volatile sulphides and gravimetrically for total S and sulphate-S. Organic C and organic N determinations were made at 600 °C to minimize breakdown of carbonates⁸. Both results agreed closely with determinations at 1,200 °C following treatment with dilute HCl which removes calcareous material, exchangeable ammonium and also some fixed ammonium⁹. Organic C analyses in both cases also agreed closely with an alternative combustion method¹⁰ and are consistent with those published elsewhere¹¹.

Water contents were measured before sampling (G. S. Boulton, personal communication) and decreased from 22.7% near the top of the core to 16.1% at the base. For porewaters of average seawater composition, marine sulphate could not exceed 0.01% S and most sulphate must arise from sulphide oxidation. Total S is accordingly treated as sulphide-S.

The depth profiles demonstrate the following differences in the amounts and compositions of the phases participating in diagenesis (Fig. 1). (1) Organic C (0.16–0.27%) and total S (0.05–0.11%) contents are both smaller in the upper unit than in the lower unit (organic C 0.34–0.42%, total S 0.26–0.41%). (2) The total sulphur, C/S and C/N ratios change progressively with depth through the upper unit and across the boundary, reaching near-constant values at depths >25 cm.

These features arise as a result of microbiological diagenesis of sedimentary organic matter, principally by sulphate reduction.



Bacterial sulphate reduction occurs only in the absence of oxygen but such conditions are common in marine sediments, which then allow anaerobic sulphate-reducing bacteria to oxidize metabolizable organic matter to dissolved carbonate species and so produce H_2S by the reduction of seawater sulphate. The H_2S

reacts with detrital iron minerals to give black, acid-volatile iron sulphides which are then slowly converted to pyrite. Sulphide contents thus increase with depth (from low surface values) in Recent sediments deposited in conditions where depth and time of burial are related. Such gradients have been described in many different depositional environments^{12–14}. Furthermore, in sediments deposited under an oxygenated water column, pyrite formation is apparently limited by the fraction of organic matter which can be metabolized. Thus continuing sulphate reduction is accompanied by a decreasing ratio of organic C to sulphide-S (C/S), which typically reaches a constant value at depths of <2 m, when metabolizable organic C is mostly exhausted^{13–15}. Total organic C is a poor measure of metabolizable organic C in a sediment, since the latter depends on the extent of preservation in the water column and is also affected by benthic activity and terrestrial contributions¹⁶. However, most organic N is present in proteins which undergo rapid preferential destruction such that the C/N mole ratios of organic matter reach constant values, typically of about 15 or more, after only 1–2 m burial in Recent sediments^{17,18}.

Recent sediments deposited in a depth–time sequence thus show near-surface gradients of an increasing sulphide content, increasing C/N ratio and a decreasing C/S ratio. These gradients can provide a useful signature of age. The observed gradients in total S, C/N and C/S ratios in the RISP sediments are similar to those observed in Recent sediments and are also internally consistent, in that the addition of sulphide (gradients in total S and C/S) corresponds to a decrease in the metabolizable fraction of the organic matter (as indicated by the C/N gradient). A Recent origin, at least for the upper unit, is also indicated by the C/N ratios, which are close to the range in diatoms and plankton (5–6.6)¹⁷. Contributions from fixed ammonium may cause these values to be low, although most fixed ammonium is associated with illitic clays¹⁹ which rarely exceed 30% of similar Ross Sea Miocene sediments²⁰. The near-surface RISP samples have C/N values which clearly indicate that the sediment was mixed with fresh organic material on deposition. Such material typically survives near-surface anaerobic microbiological decay for between a few hundred to a few thousand years, as indicated by the time needed to reach uniform C/N and C/S values^{13–17}. Diagenetic nitrogen loss is slower in oxic sediments (where it does not occur concomitantly with pyrite formation) but is still completed within <120,000 yr (ref. 19), except where

rates of bacterial metabolism may be limited by the diffusion of electron acceptors through a large thickness of sediment²¹. Thus, although the absence of benthic activity²² and low metabolic rates in the water column²³ may diminish the decay of metabolizable nitrogenous material, it seems certain that the upper unit contains organic matter of Recent (or at the very oldest, late Pleistocene) origin. Radiocarbon ages support a Recent origin⁴. Recycled spores and pollen have been identified in RISP sediments^{2,11}, but the C/N ratios of plant material (13–26) are much larger than the C/N values in the upper unit, and recycled terrestrial material cannot be the main organic matter source.

The diagenetic gradients indicate that deposition of the upper unit has occurred postglacially, with fresh organic matter being incorporated during the reworking of older sediments. Thus the geochemical data are consistent with the occurrence of a reworking event which mixed Pliocene and Pleistocene fauna with the Miocene forms⁴, but not with the existence of an *in situ* Miocene succession^{1,2}. No reworking of the upper unit has occurred following deposition at the present site, otherwise no depositional or diagenetic variations would occur. In this respect, we agree with the conclusions based on ¹⁰Be measurements, although these imply a pre-Quaternary origin at the latest for the upper unit²⁴. Such a prolonged period of non-deposition is incompatible with the existence of reduced S species, and nitrogenous organic matter, at the sediment–water interface. Comparatively large ¹⁰Be ages may result from diminished rates of ¹⁰Be deposition at high latitudes and in the presence of an ice cover²⁵.

As the two units are so similar in detrital mineralogy, their main chemical differences arise as a result of alteration by *in situ* diagenesis which differs only in degree and not in direction. The extent of alteration by diagenetic sulphate reduction is mainly controlled by metabolizable organic matter concentrations and a fundamental difference between the two units is the lower organic C content of the upper unit, even though the diagenetic loss of organic C is not yet completed here. Hence, the lighter colour of the upper unit is probably due to a smaller content of black, iron monosulphides arising from both a lower initial organic C content and unfinished sulphate reduction. Complete diagenetic alteration of the upper unit would result in a sediment with less organic C and S than the lower unit, although the progressive C/N and C/S trends would suggest that the two units would ultimately be very similar in these respects.

The total S, C/S and C/N trends appear to pass through the iron oxide layer (at 18-cm depth) without major perturbation, but by 25 cm have closely approached the values at greater depths. The depth at which diagenesis is mostly complete thus lies only a little below the iron oxide layer and hence there are no unequivocal signals of depth–age relationships in the lower unit. However, the small but irregular trends of increasing total S and C/N values would be consistent with continuity of diagenesis through the iron oxide layer, and hence with a post-late Pleistocene origin for the whole core.

The iron-stained layer between the two units is a major chemical discontinuity whose temporal significance is unclear. No samples at this depth were analysed, but the small kink in the C/S ratio at 15–16 cm suggests a possible loss of S which, together with the presence of iron oxides, implies a discontinuity in sedimentation which allowed the oxidation of sulphidic sediments by exposure to oxygenated seawater. The oxidation of sulphides may proceed quite rapidly under these circumstances and numerical models for the oxidation of sulphidic sediments by the diffusive penetration of dissolved oxygen into the RISP sediments indicate that a 1-cm oxide layer could be produced in as little as 100 yr.

The diagenetic chemistry clearly demonstrates that the RISP sequence is not wholly Miocene and that the upper unit is probably of Recent origin. Unfortunately, there are no clear indications as to the age of the lower unit. Our observations can be reconciled with the presence of Miocene, also Pliocene

and Pleistocene, fauna if the upper unit was formed by reworked material (similar to that in the lower unit) which was re-deposited following suspension in the water column (which would allow an autochthonous organic matter contribution). There are no chemical features which are inconsistent with a post late-Pleistocene origin for the whole core.

We thank the RISP support group, especially John Clough, the Office of Polar Programmes and Dr G. S. Boulton.

Received 26 October 1984; accepted 9 April 1985.

1. Brady, H. T. *Antarct. J.* **8**, 123–124 (1978).
2. Brady, H. T. & Martin, H. *Science* **203**, 437–438 (1979).
3. Webb, P. N. *RISP Tech. Rep.* 79-1 (University of Nebraska, 1979).
4. Kellogg, T. B. & Kellogg, D. E. *Nature* **293**, 130–133 (1981).
5. Brady, H. T. *Nature* **303**, 510–511 (1983).
6. Kellogg, T. B. & Kellogg, D. E. *Nature* **303**, 511–513 (1983).
7. Webb, P. N. *et al. Science* **203**, 435–437 (1979).
8. Hudson, J. D. *Geochim. cosmochim. Acta* **31**, 2364–2378 (1967).
9. Scott, A. D., Hanway, J. J. & Edwards, A. P. *Soil Sci. Soc. Am. Proc.* **22**, 388–392 (1958).
10. Krom, M. D. & Berner, R. A. *J. sedim. Petrol.* **53**, 660–663 (1983).
11. Wrenn, J. H. & Beckman, S. W. *Antarct. J.* **16**, 69–71 (1981).
12. Berner, R. A. *Am. J. Sci.* **268**, 1–23 (1970).
13. Goldhaber, M. B. & Kaplan, I. R. in *The Sea* Vol. 5 (ed. Goldberg, E. D.) 569–655 (Wiley, New York, 1974).
14. Filipec, L. H. & Owen, R. M. *Am. J. Sci.* **280**, 1097–1112 (1980).
15. Berner, R. A. *Am. J. Sci.* **282**, 451–473 (1982).
16. Demaison, G. J. & Moore, G. T. *Bull. Am. Ass. Petrol. Geol.* **64**, 1179–1209 (1980).
17. Emery, K. O. & Rittenberg, S. C. *Bull. Am. Ass. Petrol. Geol.* **36**, 735–806 (1952).
18. Rosenfield, J. K. *Am. J. Sci.* **281**, 436–462 (1981).
19. Stevenson, F. J. & Cheng, C.-N. *Geochim. cosmochim. Acta* **36**, 653–671 (1972).
20. Cook, H. E., Zimmels, I. & Matti, J. C. *Init. Rep. DSDP* **28**, 981–998 (1975).
21. Waples, D. W. & Sloan, J. R. *Geochim. cosmochim. Acta* **44**, 1463–1470 (1980).
22. Lipps, J. H., Ronan, T. E. Jr & De Luca, T. E. *Science* **203**, 447–449 (1979).
23. Azam, F. *et al. Science* **203**, 451–453 (1979).
24. Yiou, F. & Raisbeck, G. M. *EOS* **62**, 297 (1981).
25. Faure, G. & Taylor, K. C. in *Antarctic Earth Science*, 546–549 (Australian Academy of Science, Canberra, 1983).

Late-Quaternary climatic change on the American North Pacific Coast

C. J. Heusser*, L. E. Heusser† & D. M. Peteet‡

* Department of Biology, New York University, Box 608, Tuxedo, New York 10987, USA

† Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

‡ NASA/Goddard Space Flight Center, Institute for Space Studies, New York, New York 10025, USA

Temporal changes in solar radiation caused by variations in the Earth's orbit figure prominently in current Quaternary climate theory^{1–4}. Experimental design of atmospheric circulation models emphasizes the importance of solar radiation in the heating of interglacial land and ocean surfaces at different seasons of the year, whereby monsoon-type circulation—seasonal shifting of regional cyclonic and anticyclonic centres—develops, while during glacial ages, seasonality is greatly modified by the influence of ice sheets^{5–7}. We investigated the late-Quaternary climate of the North Pacific, where according to modelling, solar radiation in the early Holocene at the time of the summer solstice is high and in the late Holocene is relatively low. We compared quantitative temperature and precipitation estimates from southern Alaska, obtained by application of transfer functions to fossil pollen records, with estimates from western Washington⁸ and southern British Columbia⁹. Data extending over >10,000 years show a broadly consistent pattern of climatic change in general agreement with predicted variations in solar radiation and their effect on atmospheric circulation and seasonal duration of pressure systems over the North Pacific Ocean. In the early Holocene, as monsoon-type circulation became established with melting of glaciers, the subtropical North Pacific anticyclone annually regulated climate for a longer period at higher latitudes than at present, so that warmth and dryness (high summer radiation) increased in southern Alaska. The Aleutian low-pressure centre, locus of cyclonic storm activity, intensified during the late Holocene, resulting in colder and more humid coastal climate (low summer radiation) over much of the year and increased frequency of glacier growth in the cordillera.

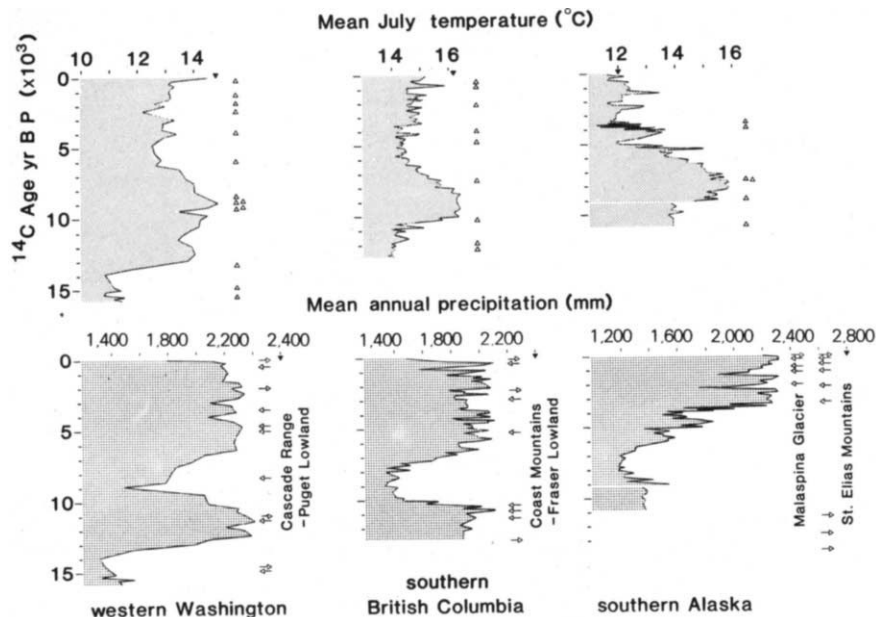


Fig. 1 Late-Quaternary mean July temperature and annual precipitation estimated by use of transfer functions for sectors of North America located in western Washington ($47^{\circ}49'N$, $124^{\circ}12'W$)⁸, southern British Columbia ($49^{\circ}15'N$, $122^{\circ}20'W$)⁹, and southern Alaska ($60^{\circ}01'N$, $141^{\circ}57'W$). Data are drawn to a uniform time scale with chronology indicated by radiocarbon ages (Δ) and assuming constant sedimentation rates between ages. Markers on temperature and precipitation scales are modern means. Arrows pointing left next to precipitation curves signify times of glacier advance, whereas retreat is indicated by arrows pointing right; vertical lines between arrows indicate extended intervals of glacier activity.

Under the influence of a trough of low pressure located in the Aleutian/Gulf of Alaska region, climate today along the American north-west coast is mostly stormy, cold and rainy¹⁰. In summer, as the subtropical anticyclone develops over the North Pacific Ocean, the Aleutian low is weakened and displaced northwestward by the build up of high pressure. Some weeks of relatively stable conditions ensue in the western conterminous United States and southwestern Canada, providing a contrast with the weather during the colder part of the year. Although summer months in southern Alaska and northern coastal British Columbia are driest, influenced by the seasonal pressure change, cyclonic storms, accompanied by heavy precipitation and cloudiness, continue to strike this sector of the coast.

Seasonal duration and extent of pressure systems over the North Pacific were variable in the late Quaternary, as implied by quantified mean July temperature and annual precipitation estimates for three geographic areas (Fig. 1). Data from southern Alaska serve for comparison with data from the Olympic Peninsula of western Washington⁸ and the Coast Mountains of southern British Columbia⁹. All three data sets were generated by two independent regression equations⁸ that relate modern composition of pollen in surface samples from the Pacific coastal forest and tundra regions to recorded temperature and precipitation values at sites between the Aleutian Islands and California. Applied to radiocarbon-dated fossil pollen records in the three areas, the equations yield estimates of late-Quaternary climatic parameters. Transfer functions used were applied within the limits of the calibration data set.

Southern Alaska data (Fig. 1) are derived from fossil pollen contained in two sections of a muskeg located near Munday Creek on the Gulf of Alaska, ~15 km north-west of Icy Cape in the Malaspina Glacier district. One section extends to ~9,000 yr BP¹¹ whereas the other reaches beyond to 10,820 yr BP¹². Mean July temperature is near $14^{\circ}C$ at the end of the Pleistocene (10,000 yr BP) and in the early Holocene, reaching a maximum close to $16^{\circ}C$ at ~8,000 yr BP and then decreasing to $12^{\circ}C$ at ~5,000 yr BP, with fluctuations around this value until the present. Annual precipitation, just under 1,400 mm at the beginning of the record, reached a minimum near 1,200 mm, coinciding with the temperature maximum at ~8,000 yr BP, and thereafter increased to almost twice as much during the past four millennia, while temperature fell to its lowest level of the entire time span.

Temperature and precipitation trends expressed by the curves represent a pattern, repeated to the south in British Columbia and Washington (Fig. 1). All sites show early Holocene temperature maxima and precipitation minima, whereas in the late

Holocene (after 5,000 yr BP), as in the late Pleistocene, higher precipitation is coincident with lower temperature. Increased warmth and dryness during the early Holocene suggest extended dominance by the subtropical Pacific anticyclone; conversely, before and after this interval, when precipitation amounts are higher and temperature values lower, polar Pacific cyclones became effective in generating storms that increased in frequency over a greater part of the year. Changes associated with the centres of atmospheric pressure were influenced to a great extent by land-ocean temperature contrasts, latitudinal temperature gradients and topography. Changes follow closely predictive general circulation models in which monsoon-type climate is related to variations in seasonal solar radiation intensity⁵⁻⁷.

Storm activity, implied by the fossil evidence, was most intense in the late Holocene in southern Alaska, where precipitation increased by some 1,200 mm and temperature fell ~ $4^{\circ}C$. Particularly wet climate effecting generation of the muskeg studied near Icy Cape is indicated by an unusually high deposition rate (42 mm yr^{-1} , assuming no compaction) for 150 cm of unhumified peat formed between 3,860 and 3,500 yr ago; before this, under a warmer, drier climate between 9,000 and 7,700 yr ago, rate of deposition of 25 cm of highly humified peat was much lower (2 mm yr^{-1}). Rapid peat deposition close to 3,500 yr ago has been found elsewhere in southern Alaska¹³.

This pronounced climate change occurred at the time of the Neoglaciacion, during which colder wetter conditions, giving rise to excessive snowfall in the lofty St Elias Mountains, caused glacier regeneration (Fig. 1) between 3,300 and 2,100 yr ago, 1,250 and 1,030 yr ago and in recent centuries, after an extended interval of early Holocene recession¹⁴⁻¹⁸. During the Neoglaciacion, ice flow from the St Elias Mountains contributed to the formation of the huge piedmont lobe of Malaspina Glacier, which spread across the coastal lowland and continental shelf to the Pacific. Episodes of ice advance in the Malaspina Glacier district occurred over the past two millennia, with maxima dated between 1,600–1,200 and 830–560 yr ago and in the past few centuries¹⁹⁻²³.

Malaspina Glacier, covering some 2,200 km², rests mostly below sea level²⁴ along a part of the Gulf of Alaska coast that, under the warmer, drier climate of the early Holocene, was probably an embayment. Located between the eastern end of the Robinson Mountains at Icy Bay and the Deception Hills at the mouth of the Alek River, the embayment extended along >200 km of coastline, including present-day Icy and Yakutat Bays, before the Neoglaciacion. Higher temperature and lower precipitation estimates make it unlikely that mass balance of Malaspina Glacier in its present lobate form could have been maintained in the early Holocene.

Greater frequency of storms passing across southern British Columbia and Washington over the past six or seven millennia is reflected by the estimates of increased precipitation in this part of the coast (Fig. 1). The increase, while temperature steadily decreased, inaugurated late Holocene glacial episodes at ~5,000 yr ago²⁵⁻³¹. During the early Holocene, except for isolated instances of minor glacial advance in the Cascade Range³²⁻³⁴, glaciers were inactive. Climate, under control of the subtropical anticyclone, was apparently too warm and dry for glacial growth. In the late Pleistocene (15,000–10,000 yr ago), a major advance occurred in the Puget Lowland in a cold, dry climate ~14,500 yr ago³⁵, followed under wetter but less cold conditions by minor advances in the Fraser Lowland and North Cascades close to 11,000 yr ago³⁶⁻³⁸ and in the Coast Mountains ~10,500 yr ago³⁹. Late Pleistocene glaciation of the Puget Lowland has been attributed to northward movement of storm tracks, causing heavy snowfall in the glacier source region in the Coast Mountains of British Columbia, coincident with decay of the Laurentide ice sheet⁴⁰.

In the long-term, late-Quaternary climate fluctuations along the American north-west coast, keyed to shifting dominance of atmospheric pressure cells over the North Pacific, are outlined by the pattern of temperature and precipitation estimates. They conform with temporal changes in seasonal radiation intensity predicted by experimental modelling⁵⁻⁷, previously brought out by fossil pollen studies in north-west Canada⁴¹. In the absence of tighter radiocarbon-dated time control over extended intervals in the records, assessment of short-term fluctuations of the estimates (of the order of centuries) is not justified. Rates of response to climate change, as well as speed of migration northward following deglaciation of the British Columbia and Alaskan coasts, by plant species used in the equations, for example, are variables that must enter into the timing of events. These may be significant, but further dating and replication are required to define them.

We thank S. C. Porter for unpublished data on glacier fluctuations in the Malaspina Glacier district and C. A. Sancetta, D. H. Rind and W. F. Ruddiman for critical comments on the manuscript. Final version of the letter was completed while C.J.H. and L.E.H. were in residence at Clare Hall, University of Cambridge. We thank N. J. Shackleton of The Godwin Laboratory and R. G. West of The Botany School for courtesies extended during our stay. Supported by NSF grant DEB-7910505 to New York University with assistance from the US Geological Survey.

Received 23 January; accepted 27 March 1985.

- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. & Imbrie, K. P. *Ice Ages, Solving the Mystery* (Enslow, Short Hills, 1979).
- Imbrie, J. & Imbrie, J. Z. *Science* **207**, 943–953 (1980).
- Gates, W. L. in *Climatic Variations and Variability: Facts and Theories* (ed. Berger, A.) 435–459 (Reidel, Dordrecht, 1980).
- Kutzbach, J. E. *Science* **214**, 59–61 (1981).
- Kutzbach, J. E. & Otto-Bliessner, B. L. *J. Atmos. Sci.* **39**, 1177–1188 (1982).
- Kutzbach, J. E. & Guetter, P. J. *Ann. Glaciol.* **5**, 85–87 (1984).
- Heusser, C. J., Heusser, L. E. & Streeter, S. S. *Nature* **286**, 702–704 (1980).
- Mathewes, R. W. & Heusser, L. E. *Can. J. Bot.* **59**, 707–710 (1981).
- Hare, F. K. & Hay, J. E. in *Climates of North America* (eds Bryson, R. A. & Hare, F. K.) 49–192 (Elsevier, Amsterdam, 1974).
- Peteet, D. M. thesis, New York Univ. (1983).
- Heusser, C. J. *Am. Geogr. Soc. Spec. Pub.* **35** (American Geographical Society, 1960).
- Heusser, C. J. in *Proc. Intern. Symp. World Climate 8000 to 0 B.C.* (ed. Sawyer, J. S.) 124–141 (Royal Meteorological Society, London, 1966).
- Denton, G. H. *Bull. geol. Soc. Am.* **85**, 871–892 (1974).
- Denton, G. H. & Stuiver, M. *Bull. geol. Soc. Am.* **78**, 485–510 (1967).
- Denton, G. H. & Karlén, W. *Quat. Res.* **3**, 155–205 (1973).
- Denton, G. H. & Karlén, W. *Quat. Res.* **7**, 63–111 (1977).
- Rampton, V. N. *Arctic* **24**, 277–300 (1971).
- Plafker, G. & Miller, D. J. *U.S. Geol. Surv. Misc. Geol. Invest. Map* 1-271 (1958).
- Plafker, G., Hudson, T., Rubin, M. & Dixon, K. *U.S. Geol. Surv. in Alaska 1980 Circ.* **844**, 111–115 (1980).
- Sharp, R. P. *Geogr. Rev.* **48**, 16–26 (1958).
- Trautman, M. A. & Willis, E. H. *Radiocarbon* **8**, 161–203 (1966).
- Molnia, B. F. *Proc. Offshore Tech. Conf. No. 9* **3**, 115–126 (1977).
- Sharp, R. P. *Bull. geol. Soc. Am.* **69**, 617–646 (1958).
- Mathewes, W. H. *J. Geol.* **59**, 357–380 (1951).
- Stuiver, M., Deevey, E. S. & Gerasinski, L. J. *Am. J. Sci. Radiocarbon Suppl.* **2**, 49–61 (1960).
- Crandell, D. R. & Miller, R. D. *U.S. Geol. Surv. Prof. Paper* 847, 1–59 (1974).
- Miller, R. D. *Arctic Alp. Res.* **1**, 49–66 (1969).
- Fulton, R. J. *Geol. Surv. Can. Paper* 71–37 (1971).
- Clague, J. J. *Geol. Surv. Can. Paper* 80–35 (1981).
- Burke, R. M. & Birkeland, P. W. in *Late-Quaternary Environments of the United States*. Vol. 2 (ed. Wright, H. E., Jr) 3–11 (University of Minnesota Press, 1983).

- Beget, J. E. *Geology* **9**, 409–413 (1981).
- Waitt, R. B., Jr, Yount, J. C. & Davis, P. T. *Quat. Res.* **17**, 191–210 (1982).
- Beget, J. E. *Quat. Res.* **21**, 304–316 (1984).
- Waitt, R. B., Jr & Thorson, R. M. in *Late-Quaternary Environments of the United States*. Vol. 1 (ed. Porter, S. C.) 53–70 (University of Minnesota Press, 1983).
- Armstrong, J. E. *Bull. geol. Soc. Am.* **67**, 1666–1667 (1956).
- Porter, S. C. *Quat. Res.* **10**, 30–41 (1978).
- Porter, S. C., Pierce, K. L. & Hamilton, T. D. in *Late-Quaternary Environments of the United States*. Vol. 1 (ed. Porter, S. C.) 71–111 (University of Minnesota Press, 1983).
- Armstrong, J. E. *Geol. Soc. Am. Progr. 1966 Ann. Meetings*, 7 (1966).
- Hicock, S. R., Hebda, R. J. & Armstrong, J. E. *Can. J. Earth Sci.* **19**, 2288–2296 (1982).
- Ritchie, J. C., Cwynar, L. C. & Spear, R. W. *Nature* **305**, 126–128 (1983).

Evidence for an early Plio-Pleistocene rainforest expansion in East Africa

P. G. Williamson

Department of Geological Sciences and
Museum of Comparative Zoology, Harvard University, Cambridge,
Massachusetts 02138, USA

The Plio-Pleistocene palaeoenvironmental history of the Turkana Basin and neighbouring areas in north Kenya and southern Ethiopia has been a matter of considerable interest due to the important palaeoanthropological, palaeoarchaeological and palaeontological discoveries that have been made in this region¹⁻³. The consensus is that although environmental oscillations have certainly occurred over the past 4 Myr, Plio-Pleistocene environments were generally comparable to the modern environment of the area⁴. However, the recent recovery of two taxa presently characteristic of the central African rainforest—the anacardiacean tree *Antrocaryon* and the prosobranch gastropod *Potadoma*—from a restricted chronostratigraphic horizon dating at ~3.3–3.4 Myr, at widely separated sites in the Turkana region, suggests that the general pattern of environmental stability was punctuated at this time by a brief but significant rainforest extension in East Africa. This brief humid interval apparently coincides with certain important molluscan immigrations, and can be broadly correlated with the first major episode of climatic deterioration in the Northern Hemisphere.

Recent reviews of Plio-Pleistocene palaeoenvironments in the Omo Valley and Koobi Fora areas of the Turkana Basin, north Kenya, have drawn on sedimentological, palaeontological and palaeobotanical evidence (see, for example, refs 1–4). Although a relatively xeric interval between 2.5 and 2.0 Myr has been postulated⁵, the evidence for significant environmental change at this time has been questioned⁴. The general consensus seems to be that throughout the past 4 Myr the Turkana region has been characterized by a suite of terrestrial environments similar to those found at present—dry tropical scrub and grasslands of the ‘Sudan savanna’ replaced locally by limited gallery forest in the neighbourhood of perennial drainages⁴. However, several recent lines of evidence suggest that the general pattern of overall environmental stability in the latest Cenozoic of East Africa may have been briefly but significantly perturbed by an important humid episode at 3.3–3.4 Myr, an episode which entailed a major expansion of the central African rainforest. The suggestion that a humid episode occurred at this time was originally prompted by the discovery of fruits of the anacardiacean rainforest genus *Antrocaryon*⁶ in the ‘Brown Sands’ locality of the Usno Formation, a short stratigraphic distance above tuff U10, and from Member A1.1 of the Shungura Formation in the lower Omo Valley (see Fig. 1), immediately underlain by tuff A and overlain by tuff B. Tuff U10 is considered to be a lateral correlate of tuff B of the Shungura Formation (F. H. Brown, personal communication), and tuff B itself is now considered a lateral correlate of the Tulu Bor Tuff from the Koobi Fora area^{7,8}. The age of this latter horizon is considered to be ~3.35 Myr, on the basis of its stratigraphic position between the immediately overlying Toroto Tuff (firmly dated at 3.32±0.02 Myr) and the underlying Lokochot Tuff⁹. On the basis of geochemical comparisons, the Lokochot has been correlated with tuff A of the Shungura, which is known to lie close to the boundary between

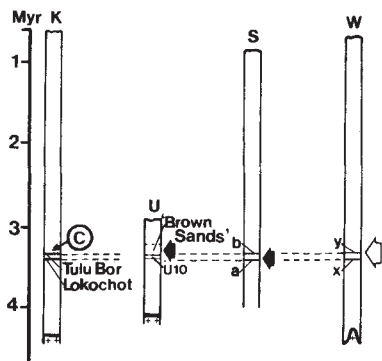


Fig. 1 Stratigraphic distribution of *Antrocaryon* (solid arrows) in the Usno (U) and Shungura (S) formations of the Omo Valley; occurrence of *Potadoma* (open arrow) in sections west of Lake Turkana (W) and the earliest African occurrence of *Corbicula* (C), in the Koobi Fora area (K). a, Tuff A of the Shungura Formation, which can be correlated with the Lokochot Tuff of the Koobi Fora area and also with a tuff (indicated here by x) in the areas west of Lake Turkana; b indicates tuff B of the Shungura Formation, which can be correlated with Tuff U10 of the Usno Formation, with the Tulu Bor Tuff of the Koobi Fora Formation, and with a tuff (indicated here by y) in the area west of Lake Turkana. Note the extremely restricted stratigraphic distribution of both *Potadoma* and *Antrocaryon*, and the fact that their occurrence coincides closely with the first African record of *Corbicula*. Data from various sources. The geographical distribution of these occurrences is indicated in Fig. 2. General chronostratigraphic framework after Brown *et al.*⁹

the Gauss and Gilbert chrons; the Lokochot/tuff A horizon must accordingly date at ~3.42 Myr⁹. *Antrocaryon* in the lower Omo Valley is therefore confined to an interval from ~3.3 Myr to ~3.4 Myr. In addition to the occurrence of *Antrocaryon* in the Omo Valley deposits, the striking development of palaeosols in Member B has been considered evidence for humid conditions¹⁰, and the mammal fauna from Member B10 contains rodents and prosimians (notably *Galago*) which have been interpreted as indicating a rainforest setting and forest connections with central Africa¹¹. However, the various sites from which this evidence was obtained cover a relatively small area of the Omo Valley and all are fluvial deposits in close proximity to the presumed axis of the original Omo drainage⁴. Consequently, the evidence for rainforest conditions around the tuff B level could simply reflect a restricted local development of riverine gallery forest rather than a regionally significant rainforest expansion, as Brown⁴ and others have pointed out⁶.

The suggestion that a regionally significant humid episode and concomitant rainforest expansion occurred at 3.3–3.4 Myr in northern Kenya and southern Ethiopia is greatly strengthened by the recent discovery of the rainforest prosobranch *Potadoma*¹² from two faunas, separated stratigraphically by some 2 m, in the Lomekwi drainage area west of Lake Turkana some 60 km south of the Omo Valley (see Fig. 2). These faunas occur at the level of a tuff which has been correlated, on geochemical grounds, with the Tulu Bor (U10; tuff B) horizon⁹ (see Fig. 1); this occurrence of *Potadoma* constitutes the only definite record of this west and central African genus in East Africa¹².

Both *Antrocaryon* and *Potadoma* are at present restricted to the central African rainforest^{5,12,13}; indeed, their joint distribution essentially delimits the present distribution of rainforest in central Africa (see Fig. 3). The simultaneous occurrence of these distinctive taxa in sections up to 150 km apart in the Turkana region at 3.3–3.4 Myr strongly suggests, therefore, a dramatic expansion of humid rainforest conditions in the area at this time. However, the extremely restricted chronostratigraphic distribution of these taxa in the Turkana area deposits suggests that this period of enhanced humidity was relatively brief. The regional nature of this climatic change may also be reflected by the fact that a fauna some 8 m above the Tulu Bor Tuff in collection area 129 of the Koobi Fora area has recently provided

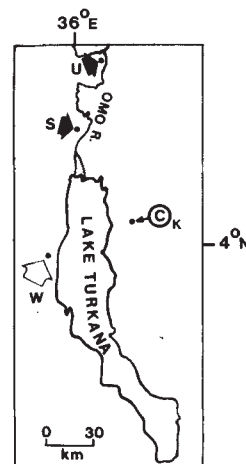


Fig. 2 Geographical location of occurrence of *Antrocaryon* (solid arrows) in the Usno (U) and Shungura (S) formations of the Omo Valley; *Potadoma* (open arrow) in the area of the Lomekwi drainage (W), west of Lake Turkana, and the first African occurrence of *Corbicula* C in area 129 of the Koobi Fora area (K) east of Lake Turkana.



Fig. 3 Present distribution of *Antrocaryon* and *Potadoma* in Africa. Black areas indicate present co-occurrence of *Antrocaryon* and *Potadoma*, except the small isolated area to the north, in which *Potadoma* occurs alone. Grey (stippled) areas indicate the present distribution of *Antrocaryon*. The circled area (T) indicates the location of the Turkana Basin. Distribution of *Antrocaryon* after Bonnefille and Letouzey⁶; distribution of *Potadoma* after Brown¹³.

the first African record of the primarily Eurasian bivalve *Corbicula*¹⁴. *Corbicula* has a long Cenozoic record in southern Eurasia, but is a comparatively recent addition to the African malacofauna¹⁵. *Corbicula* is unknown from any African malacofaunas definitely pre-dating its occurrence in area 129; it is absent from faunas lower in the Koobi Fora Formation¹⁴, from the lower Usno or Mursi formations or the Basal Member of the Shungura in the Omo Valley¹⁶, from the Lothagam or Kanapoi formations in the southern Turkana area¹⁶, and from deposits such as the Kaiso Formation in the western rift¹⁵. However, the genus predominates in the majority of faunas in the Koobi Fora Formation subsequent to its first appearance above the Tulu Bor level in area 129 (ref. 14), and *Corbicula* is now one of the most abundant and widespread African freshwater bivalves¹⁸. A regional episode of enhanced humidity, as suggested by the occurrence of *Antrocaryon* and *Potadoma*, may well have encouraged the immigration of *Corbicula* from southern Eurasia by enhancing hydrographical connections between the latter area and northeastern Africa. However, it should be noted that in area 129 the Tulu Bor is overlain by a disconformity⁹, and the first occurrence of *Corbicula* may therefore post-date the Tulu Bor by a greater interval than that indicated in Fig. 1, although no evidence of disconformity is apparent between the upper surface of the Tulu Bor and the fauna yielding *Corbicula*. Any molluscan interchange between north-east Africa and southern Eurasia is unlikely to have been strictly 'one-way', and it is therefore of interest that the only known Eurasian occurrence of the African prosobranch *Bellamyia unicolor* (Olivier) is from the 'Erq el-Ahmar beds of the

central Jordan Valley¹⁹; these beds are not directly dated but immediately overlie a basalt (the 'cover basalt') recently dated at ~3.2 Myr²⁴.

A major humid episode and consequent expansion of rainforest in East Africa at 3.3–3.4 Myr should be reflected by palaeoclimatic events elsewhere. In fact, the regional episode of rainforest expansion discussed here appears to correspond broadly to the early Gauss 'ice-age transition' episode, long considered to signify the initial formation of significant Northern Hemisphere ice sheets^{20–23}. However, this transition episode essentially represents a unidirectional palaeoclimatic deterioration reflecting the onset of ice-age conditions²³, whereas the extremely restricted chronostratigraphic distribution of *Antrocaryon* and *Potadoma* in East Africa suggests a palaeoclimatic 'punctuation' or 'overshoot' of relatively brief duration.

I thank F. H. Brown, who originally located the *Potadoma* material and provided stimulating discussions as to its significance, and D. S. Brown and J. P. Kennett for valuable comments.

Received 19 February; accepted 27 March 1985.

1. Coppens, Y., Clark Howell, F., Isaac, G. L. & Leakey, R. E. F. (eds) *Earliest Man and Environments in the Lake Rudolf Basin* (University of Chicago Press, 1976).
2. Findlater, L. C. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) (Scottish Academic, Edinburgh, 1978).

3. Leakey, R. E. F. & Leakey, M. (eds) *Koobi Fora Res. Project Monogr. 1* (Clarendon, Oxford, 1978).
4. Brown, F. H. *Am. Ass. Adv. Sci. Spec. Monogr.* 158 (1981).
5. Bonnefille, R. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y., Clark Howell, F., Isaac, G. L. & Leakey, R. E. F.) (University of Chicago Press, 1976).
6. Bonnefille, R. & Letouzey, R. *Adansonia* 16, 65–82 (1976).
7. Brown, F. H. & Cerling, T. E. *Nature* 299, 212–215 (1982).
8. Cerling, T. E. & Brown, F. H. *Nature* 299, 216–221 (1982).
9. Brown, F. H., McDougall, I., Davies, T. & Maier, R. *Palaeoanthropology: The Hard Evidence* (Liss, New York, in the press).
10. de Heinzelin, J. & Haesaerts, P. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y., Clark Howell, F., Isaac, G. L. & Leakey, R. E. F.) (University of Chicago Press, 1976).
11. Jaeger, J. J. & Wesselman, H. B. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y., Clark Howell, F., Isaac, G. L. & Leakey, R. E. F.) (University of Chicago Press, 1976).
12. Williamson, P. G. *J. Conchol.* (in the press).
13. Brown, D. S. *Freshwater Snails of Africa and Their Medical Importance* (Taylor & Francis, London, 1980).
14. Williamson, P. G. thesis, Univ. Bristol (1980).
15. Gautier, A. *Annls Mus. r. Afr. cent. Sér. 800.* 67, 1–144 (1970).
16. Van Damme, D. thesis, Rijksuniv. Gent (1976).
17. Williamson, P. G. *Nature* 293, 437–443 (1981).
18. Mandahl Barth, G. *Annls Mus. r. Congo belge Sér. 800* 32, 1–206 (1954).
19. Tchernov, E. *The Early Pleistocene Molluscs of Erq el-Ahmar* (Israel Academy of Sciences and Humanities, Jerusalem, 1975).
20. Shackleton, N. J. & Opdyke, N. D. *Nature* 270, 216–219 (1977).
21. Berggren, W. A. *Init. Rep. DSDP* 12, 723–730 (1972).
22. Thunell, R. C. *Mar. Micropalaeont.* 4, 1–173 (1979).
23. Kennett, J. P. *Marine Geology* (Prentice-Hall, Englewood Cliffs, 1982).
24. Mor, D. & Steinitz, G. *Interim Rep. Geol. Surv. Israel* No. ME682 (1982).

Reconstruction of prehistoric plant production and cooking practices by a new isotopic method

Christine A. Hastorf*† & Michael J. DeNiro†

Departments of Anthropology* and Earth and Space Sciences and Archaeology Program†, University of California, Los Angeles, California 90024, USA

As archaeological research has focused increasingly on economic questions, archaeologists have collected data that reflect production and consumption of food. Macrobotanical remains, recovered by methods such as water flotation, often provide the most available and relevant data concerning production in archaeological contexts. However, these data do not necessarily reflect the proportion of crops that were consumed¹. Here we present a new method, based on isotopic analysis of burnt organic matter, allowing the characterization of previously unidentifiable plant remains extracted from archaeological contexts. We used this method to reconstruct prehistoric production, preparation and consumption of plant foods, as well as the use of ceramic vessels, in the Upper Mantaro Valley region of the central Peruvian Andes². The method will be of use to archaeologists studying these prehistoric activities in other areas of the world.

The basis for the isotopic method used here can be summarized as follows. Nitrogen isotope ratios separate modern plants into two groups: legumes, capable of fixing atmospheric nitrogen, and non-leguminous plants, which must use nitrogen from the soil^{3,4}. Carbon isotope ratios also separate modern plants into two groups, C₃ plants and C₄ and crassulacean acid metabolism (CAM) plants, which differ in the enzymes that catalyse initial CO₂ fixation^{5–7}. Because all legumes are C₃ plants and no C₄ or CAM plant is known to fix atmospheric nitrogen, combined C and N isotope analysis can be used to separate living plants into three groups: legumes, non-leguminous C₃ plants, or C₄ and CAM plants^{3–7}. Prehistorically burnt domesticated plants, identified to the species level and assigned to one of the three plant groups on the basis of this taxonomic information^{3–8}, retain their isotopic identities during burning and fossilization (Fig. 1a). Thus, isotope ratios can be used to separate prehistoric burnt plant remains into the same three groups as modern plants.

To gain information on crop production, we used the isotopic method described above to classify prehistoric plant fragments, extracted from the soil by flotation techniques, that could not be identified by morphological criteria because they had been burned so extensively. These specimens came from three sequential occupation phases in the Upper Mantaro Valley region, namely the Early Intermediate Period/Middle Horizon (EIP/MH, 200 BC to AD 1000), the Huanca I period (H-I, AD 1000–1200) and the Huanca II period (H-II, AD 1200–1470)². The identification of the three phases is based on changing settlement patterns, ceramic type frequencies and radiocarbon dates^{9,10}. Results of the isotopic identification of the morphologically unidentifiable plant remains (Fig. 1b) were added to the much larger database of plant specimens identified by morphological analysis¹⁰. From these data we determined the changes that occurred in plant production during the three occupation phases in the Upper Mantaro Valley region (Table 1). (Because we have focused on plant production and consumption, we have included only the major agricultural crops of the region in constructing the table.) Among the indigenous crops, lupines (*Lupinus mutabilis*) and beans (*Phaseolus vulgaris*) are legumes, maize (*Zea mays*) is the only C₄ plant (there are no CAM plants) and the remaining plants are all non-leguminous C₃ plants—namely quinoa (*Chenopodium quinoa*) and the tubers, which include potatoes (*Solanum tuberosum*, *Solanum juzepczukii*), oca (*Oxalis tuberosa*), ulluco (*Ullucus tuberosus*) and mashua (*Tropaeolum tuberosum*)¹⁰. The data in Table 1 indicate that during each phase, legumes and C₄ plants were produced in roughly equal amounts, but non-leguminous C₃ plants predominated as the major crops. During the H-I period there was also increased production of C₄ plants (maize) and legumes compared with the preceding and subsequent periods.

Table 1 Domestic plant group frequencies for three sequential occupation phases of the Upper Mantaro Valley

Period	Plant group frequencies (%)			No. soil samples analysed
	Non-leguminous C ₃	C ₄	Legumes	
EIP/MH	74	29	23	171
H-I	83	43	43	175
H-II	69	13	20	265

Values for each period represent the percentage of soil samples weighing 6 kg that contained the indicated plant group. This measurement is based on presence/absence of specimens in each sample and not on internal density of the plant type. The frequencies^{11–13} do not total 100% but show the relative presence of plants. The proveniences of the sites are given in ref. 10.

* Present address: Department of Anthropology and Center for Ancient Studies, University of Minnesota, Minneapolis, Minnesota 55455, USA.

To study prehistoric cooking practices and associated vessel function, we determined the isotope ratios of the organic matter encrusted on the inside surfaces of ceramic vessel shards excavated from habitations in the Upper Mantaro Valley region. Only the plainest self-slipped jar type had encrustations, although these encrustations did not occur on every self-slipped shard. From vessel reconstructions, we found that the encrustations covered the bottoms but not the upper walls of the vessels.

All shards with visible encrustations were scraped to collect the organic matter. The 71 shards whose encrustations were analysed isotopically were chosen at random from all encrusted shards collected at the five sites spanning the three time periods for which we have the production data described in Table 1.

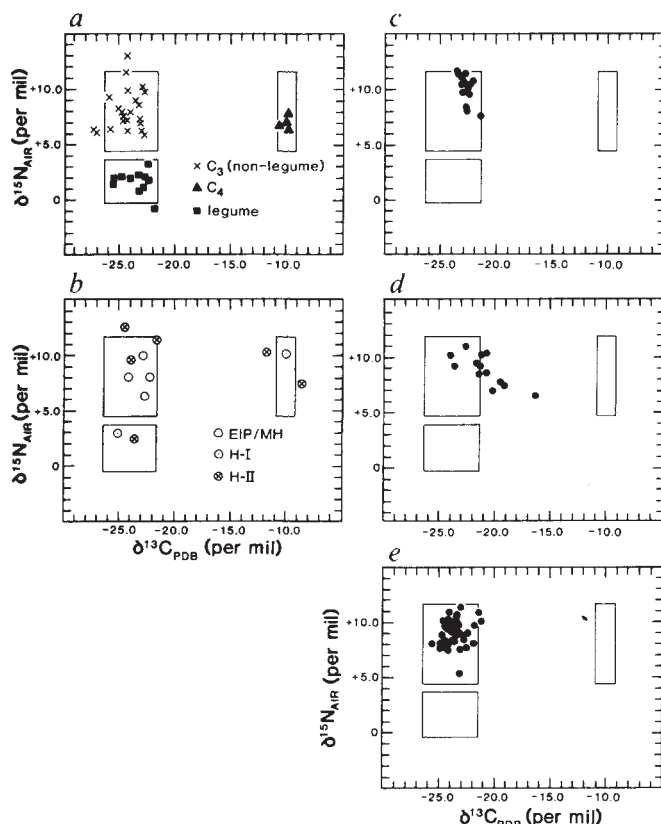


Fig. 1 C and N isotope ratios of burnt plants and of encrustations from the insides of ceramic vessel shards excavated from archaeological sites representing three continuous phases of occupation in the Upper Mantaro Valley area. The $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values are given for: *a*, prehistoric burnt plant fragments assigned to the indicated groups based on morphological criteria; *b*, prehistoric plant fragments too extensively burnt to allow such identification; *c-e*, vessel shard encrustations from the Early Intermediate Period/Middle Horizon (*c*), the H-I period (*d*) and the H-II period (*e*). Plant fragments and shard encrustations were cleaned in an ultrasonic bath and freeze-dried before the determination of their isotope ratios by methods described in ref. 8. The identified prehistoric plants in *a* came from the three occupation phases and include all the local indigenous Andean crops discussed in the text⁸. The boxes in each graph indicate the means ± 2 s.d. of the $\delta^{15}\text{N}$ values for the identified legumes and non-leguminous plants and of the $\delta^{13}\text{C}$ values for the identified C_3 and C_4 plants shown in *a*. Isotope ratios of modern plants in the three groups⁸ are similar to those indicated by the boxes. Isotope ratios are given as δ^*X values, where

$$\delta^*X = \left[\frac{(*X/X)_{\text{sample}}}{(*X/X)_{\text{standard}}} - 1 \right] \times 1,000 \text{ per mil}$$

For $\delta^{13}\text{C}$ values, $*X/X = {}^{13}\text{C}/{}^{12}\text{C}$ and the standard is the PDB carbonate (C_{PDB}). For $\delta^{15}\text{N}$ values, $*X/X = {}^{15}\text{N}/{}^{14}\text{N}$ and the standard is atmospheric nitrogen (N_{AIR}). The precisions for determination of both the $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values are ± 0.2 per mil.

For the purposes of this analysis, we assume that our random sample is representative of the total population of encrusted shards.

Analysis of the encrusted organic matter by light and scanning electron microscopy (SEM) indicates that all morphological characters useful for identification had been destroyed during burning. Three lines of evidence indicate that these organic encrustations originated from plant material that was boiled and mashed before charring. First, SEM analysis of the encrustations did not reveal the presence of bone fragments, which would have survived the burning process. Second, the atomic C/N ratios of the prehistoric encrustations were similar to those of encrustations we made by cooking and then charring modern food plants from the Upper Mantaro Valley region onto ceramic vessels, but higher than those we produced by the same process using fresh meat. The ranges and means ± 1 s.d. of the C/N ratios were as follows: prehistoric encrustations, $8.2\text{--}19.3/12.8 \pm 2.1$ ($n = 71$); modern plant encrustations, $8.2\text{--}38.8/23.4 \pm 10.0$ ($n = 8$); modern meat encrustations, $4.4\text{--}6.1/5.3 \pm 0.8$ ($n = 6$), where n is the sample size. Third, all of the prehistoric encrustations analysed were evenly distributed on the inside surfaces of the shards. The only way we could reproduce this type of distribution was by boiling and mashing fresh plants before burning them onto ceramic vessels. Dry-toasting before charring produced a spotty distribution of burnt material on the vessel. Experiments with modern plants demonstrate that isotope ratios are not substantially altered when plants are cooked and then burned onto ceramic vessels. Thus, isotope ratios of the shard encrustations indicate the type of plants cooked in these vessels.

The isotopic data for shard encrustations from the EIP/MH period (Fig. 1*c*) and the H-II period (Fig. 1*e*) show a strong clustering in the C_3 plant group box. For the intervening H-I period (Fig. 1*d*), however, the isotopic data provide evidence of C_4 plants mixed with C_3 plants in some of the encrusted samples. With the possible exception of a single sample from the H-II period (Fig. 1*e*), none of the shard encrustations shows isotopic evidence of legumes having been cooked in significant amounts in these vessels.

The ceramic shards with encrustations came from large self-slipped jars, from midden and floor excavation proveniences, suggesting that they were fragments of cooking vessels. The results from the isotopic analysis of these body shards support the hypothesis that these large jars were used for daily domestic food preparation. From this, we assume that food preparation practices reflect food consumption.

We conclude that the prehistoric inhabitants of the Upper Mantaro Valley region used these ceramic jars to cook C_3 plants alone or C_4 plants and C_3 plants together, either simultaneously or sequentially, but did not cook legumes, either alone or in combination with non-leguminous plants. These conclusions are consistent with modern cooking practices in this part of the Andes (C.A.H., unpublished). Today C_3 plants (*quinoa* or tubers) are often boiled alone for a meal, whereas when maize is boiled, it is generally prepared with C_3 plants and spices. Comparison of modern and prehistoric processing of legumes is complicated by the present-day use of fava beans (*Vicia fava*) and peas (*Pisum* sp.), which have been introduced from Europe¹⁰. Fava beans and peas are boiled as well as dry-toasted at present, whereas the indigenous legume *Lupinus mutabilis* is a casual food and tends to be toasted or eaten raw. *Phaseolus vulgaris*, although indigenous, is not in common use today. Although we do not know whether introduction of new plant foods influenced traditional cooking practices, our isotopic analysis of the shard encrustations indicates that present-day cooking patterns for indigenous plants also existed in the past.

Comparison of the consumption data (Fig. 1*c-e*) with the production data (Table 1) indicates that temporal changes in food preparation paralleled changes in food production for the C_3 and C_4 plant groups. The absence of parallel trends in the two activities for legumes and non-leguminous plants is probably related to how indigenous legumes were prepared. The evidence for maize on some H-I period shards (Fig. 1*d*), in association

with evidence of increased maize production at that time (Table 1), suggests that the additional maize available to the population during this period was eaten as a bulk carbohydrate. We suspect that during all three periods maize was being consumed as *chicha* (fermented beverage) and *cancha* (toasted corn). Because there is no evidence for significant differences in social status during the H-I period¹⁰, differential access to resources based on social position cannot have helped to produce evidence for increased consumption of maize during this period.

The isotopic method described here can be used to characterize plant materials burned onto ceramic vessels (whether the vessels were everyday cooking wares, storage jars or ceremonial items) in areas that have plants from more than one of the three isotopically identifiable plant groups. This method should also be applicable to organic encrustations on vessel exteriors, identifying plant groups used prehistorically as fuel.

We thank Henry Ajie and David Winter for technical assistance, Stanley Ambrose, Timothy Earle, Leonel Sternberg and Herb Wright for comments and the NSF (grants BNS 79-24756, ATM 79-24581, BNS 82-03723 and BNS 84-18280) and the American Association of University Women for financial support.

Received 1 October 1984; accepted 12 March 1985.

1. Dennell, R. W. *J. archaeol. Sci.* **3**, 229 (1976).
2. Earle, T., D'Altroy, T., LeBlanc, C., Hastorf, C. & Levine, T. *J. New Wild Archaeol.* **4**, 1-49 (1980).
3. Delwiche, C. C. & Steyn, P. L. *Envir. Sci. Technol.* **4**, 929-935 (1970).
4. Virginia, R. A. & Delwiche, C. C. *Oecologia* **54**, 317-325 (1982).
5. Bender, M. M. *Phytochemistry* **10**, 1239-1244 (1971).
6. Smith, B. N. & Epstein, S. *Pl. Physiol.* **47**, 380-384 (1971).
7. Bender, M. M., Rouhani, I., Vines, H. M. & Black, C. C. *Jr Pl. Physiol.* **52**, 427-430 (1973).
8. DeNiro, M. J. & Hastorf, C. A. *Geochim. cosmochim. Acta* **49**, 97-115 (1985).
9. LeBlanc, C. J. thesis, Univ. California, Los Angeles (1981).
10. Hastorf, C. A. thesis, Univ. California, Los Angeles (1983).
11. Willcox, G. A. *Anatolian Stud.* **24**, 117-133 (1974).
12. Hubbard, R. N. L. B. *Bull. Inst. Archaeol.* **12**, 197-205 (1975).
13. Hubbard, R. N. L. B. *Econ. Bot.* **34**, 51-67 (1980).

Population dynamics of *Schistosoma mansoni* in mice repeatedly exposed to infection

J. A. Crombie & R. M. Anderson

Department of Pure and Applied Biology,
Imperial College of Science and Technology, Prince Consort Road,
London SW7 2BB, UK

Studies of host resistance to parasite infection are usually based on experimental designs involving a primary infection and subsequent challenge exposure, resistance being recorded as the percentage reduction in parasite establishment in challenged hosts when compared with that in uninfected animals¹⁻⁸. Few studies have focused on the dynamic nature of helminth establishment and mortality (and their presumed dependency on the rate of current exposure and past experiences of infection) in hosts repeatedly exposed to low levels of infection⁹⁻¹¹. Here, we report the results of population studies on the dynamics of resistance to *Schistosoma mansoni* infection (a helminth parasite) in mice repeatedly exposed to cercarial invasion. Parasite burdens created by different levels and durations of exposure to infection reflect a dynamic interplay between rates of helminth establishment and mortality. Depending on the intensity of exposure, changes in worm load with duration of host infection vary from monotonic growth to a stable average parasite burden to convex curves in which the average load attains a maximum value before decaying in old animals. These trends are similar to observed patterns of *S. mansoni* infection in human communities¹²⁻¹⁴.

Under conditions of constant exposure to infection and constant worm mortality, and in the absence of acquired resistance to infection, average helminth burdens per host will rise

monotonically as the duration of exposure increases to a stable equilibrium (an 'immigration-death process')¹⁵. Patterns of change in the intensity of *S. mansoni* infection in human communities with host age, however, often show a rise to a maximum value in teenage or young adult age groups and a decline thereafter¹²⁻¹⁶. It is not known whether the observed decline in older age groups reflects age dependency in the rate of exposure to infection (that is, an ecological process) or acquired resistance to worm establishment (an immunological process)^{16,17}. We used a laboratory model to generate age-intensity curves of *S. mansoni* infection (a Puerto Rican strain) in inbred mice (CBA/Ca) by repeated exposure to constant numbers of cercariae, using four levels of cercarial exposure (Fig. 1a, b). Separate experiments revealed that ~50% of the cercariae to which a naive host was exposed became established as mature adult worms (5-week maturation period) (Fig. 1d).

The severity of the initial rise in mean worm burden with age of the host and the maximum worm burden attained are dependent on the density of cercariae to which the host is exposed. At the lowest infection dose, mice survived infection well and the experiment ran for 50 weeks with worm burdens attaining a stable plateau of ~4-6 parasites per mouse. Burdens of this magnitude may produce disease that most closely resembles human hepatosplenic schistosomiasis¹⁸. At higher exposure levels some host mortality occurred (Fig. 1a, b). In the two highest exposure experiments there was a decline in parasite burden in older mice, the severity of which was not accounted for by a dependence between host death rate and parasite burden.

As the rate of exposure of the host to infective stages was kept constant with time, these observations suggest that the acquisition of resistance to infection occurs as the duration and intensity of exposure increases. The convex pattern (Fig. 1a) recorded for the highest exposure rate could have arisen as a consequence of a time-delayed density-dependent mechanism created by limitation of resources in the host, as opposed to host defence mechanisms. This seems to be unlikely for two reasons. First, there is clear evidence for immunological response to *S. mansoni* infections⁸. Second, our experiments generated comparatively low burdens of mature and juvenile worms when compared with the much higher worm loads that can be harboured in the mesenteric veins⁸.

Perfusion of infected mice led to the recovery of juvenile worms throughout all the experiments, which suggests that parasites can become established in hosts despite long durations of past experience of infection and the presence of mature adult worms. The observed patterns of infection represent a dynamic interplay between parasite establishment (at a rate $\bar{\Lambda}$) and worm mortality (at a rate per parasite of μ). The rate of change of the mean worm burden per mouse, $M(t)$ with respect to time can be described by an immigration-death process where

$$\frac{dM(t)}{dt} = \bar{\Lambda}(\bar{M}(t)) - \mu M(t) \quad (1)$$

In agreement with current understanding of the mechanism of immunity to *S. mansoni* infection in mice⁸, we assume that acquired resistance acts to decrease the rate of parasite establishment, $\bar{\Lambda}(\bar{M}(t))$, in a manner dependent on the accumulated experience of past infection, $\bar{M}(t)$, where $\bar{M}(t) = \int_0^t M(t') dt'$.

For simplicity we assume a linear dependency where

$$\bar{\Lambda}(\bar{M}(t)) = \Lambda \left[1 - \varepsilon \int_0^t M(t') e^{-\sigma(t-t')} dt' \right] \quad (2)$$

with the constant ε denoting the efficacy of acquired resistance, described in more detail in the accompanying paper¹⁹. We further assume that the 'immunological memory' of past experience of infection is finite and decays as the time interval between the present and past experience increases with an average memory duration of $1/\sigma$ (ref. 19). The model contains four adjustable parameters, Λ , μ , ε and σ . Estimates are available for Λ and μ (see Fig. 1c, d) on the basis of our experimental design, and a crude estimate can be obtained for ε (by noting

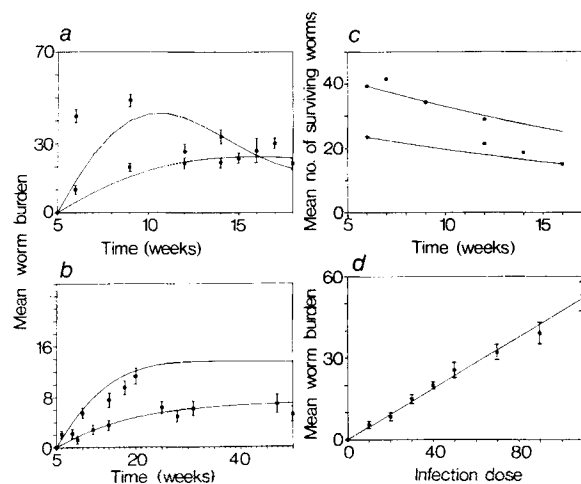


Fig. 1 *a, b*, Worm recovery through time from trickle infection experiments. Groups of 10 mice were exposed to infection at weekly intervals by paddling for 0.5 h in tanks containing a suspension of cercariae (of age <2.5 h) in 200 m dechlorinated tap water at 25 °C. Four infection densities were used: 10, 30, 100 and 300 cercariae per group per week. Worms were recovered at intervals post-infection by perfusion of the hepatic portal system with citrated saline and by general dissection. Symbols denote the mean worm recovery with standard error bars. Note that the abscissae start at 5 weeks because of the maturation delay (~5 weeks) of *S. mansoni* in laboratory mice. Solid line, predicted worm recovery calculated from the immigration-death model described in the text (equation (1)). For $(\mu - \sigma)^2 > 4\epsilon\Lambda$ the solution of equation (1) is given by

$$M(t) = A + \frac{\Lambda}{\lambda} \left[\frac{(p_1 + \sigma)e^{p_1 t}}{p_1} - \frac{(p_2 + \sigma)e^{p_2 t}}{p_2} \right]$$

where $A = \Lambda\sigma/(\sigma\mu + \epsilon\Lambda)$; $p_1 = [-(\mu + \sigma) + \lambda]/2$, $p_2 = -(\mu + \sigma + \lambda)/2$ and $\lambda = [(\mu - \sigma)^2 - 4\epsilon\Lambda]^{1/2}$. When $4\epsilon\Lambda > (\mu - \sigma)^2$ the solution adopts the form

$$M(t) = A + e^{-(\mu + \sigma)t/2} \left[\frac{2\Lambda}{\theta} \sin\left(\frac{t\theta}{2}\right) \left[1 - A\left(\frac{\mu + \sigma}{2\Lambda}\right) \right] - A \cos\left(\frac{t\theta}{2}\right) \right]$$

where $\theta = [4\epsilon\Lambda - (\mu - \sigma)^2]^{1/2}$. Equation (2) is a crude mimic of current understanding of acquired resistance mechanisms, as the value of ϵ must be small to prevent negative infection rates. The choice of this function was determined by analytical convenience and the observation that establishment decays slowly as the summed past experience of infection rises (negative infection rates prevented by a very small ϵ value). If the net infection rate in equation (1) < 0 (that is, ϵ large), acquired immunity is assumed to influence parasite establishment and survival. *a*, Record of the results for cercarial exposure levels of 10 and 30 infective stages per week (model parameter values $\epsilon = 0.0085$ per week, $\sigma = 0.2$ per week and $\mu = 0.045$ per week (see *c*). The values of Λ calculated by means of the relationship shown in *d* were 4.711 and 14.13 per week, respectively. *b*, Similar to *a* but the results for cercarial exposure levels of 1 and 3 infective stages per week ($\Lambda = 0.4711$ and 1.413 per week, respectively; other parameter values as defined for *a*). *c*, Adult worm survival over a 10-week period post-maturation. Two examples are shown where initial adult worm densities at week 6 of infection were ~40 and 25 parasites per mouse. The symbols denote observed means (from 10 mice) and the solid lines denote the fit of an exponential decay survival model with constant death rate $\mu = 0.045$ per week. *d*, Worm recovery 6 weeks post-infection of naive mice exposed to a single infection of varying intensity (0–110 cercariae). Symbols denote observed mean recoveries (10 mice) (the underlying distribution of worm numbers per inbred mouse was Poisson in form) and the solid line is the best-fit linear model with slope $b = 0.4711$ (that is, 40% of invading cercariae established within the host at week 6) where $r^2 = 0.99$. Vertical bars denote 95% confidence limits. The fit of the model to observed trends is poor in the early stages (weeks 6 and 9) of the cercariae/mouse/week experiment (*a*) as a consequence of higher than average established rates in these batches of experimental animals (higher than the average value of 47% of exposure dose recorded in *d*).

the decline in juvenile worm establishment as a function of accumulated past experience of infection). At present nothing is known about the duration of immunological memory, but model predictions crudely mimic observed changes in parasite burden as the duration of exposure to infection increases, with a fixed value of $1/\sigma = 5$ weeks (obtained by trial and error) (Fig. 1*a, b*).

The model (with no adjustable parameters other than that (Λ) set by the experimental design) is able to reflect accurately monotonic growth to a stable parasite burden for low exposures to infection and convex curves for high rates of exposure. This provides tentative support for the view that convex age-intensity curves of *S. mansoni* infection in human communities may arise partly as a consequence of acquired resistance. Observed patterns in man are probably generated by a combination of ecological and immunological processes²⁰.

To explore the significance of immunological memory, curative chemotherapy was used in conjunction with repeated exposure to infection. For low and medium levels of exposure (1 and 10 cercariae per week), there were no significant differences following drug treatment in mice with 9, 12 and 15 weeks previous experience of continual infection, when compared with primary trickle infections (Fig. 2*a, b, c, d*). Resistance following chemotherapy has been the subject of previous studies but the

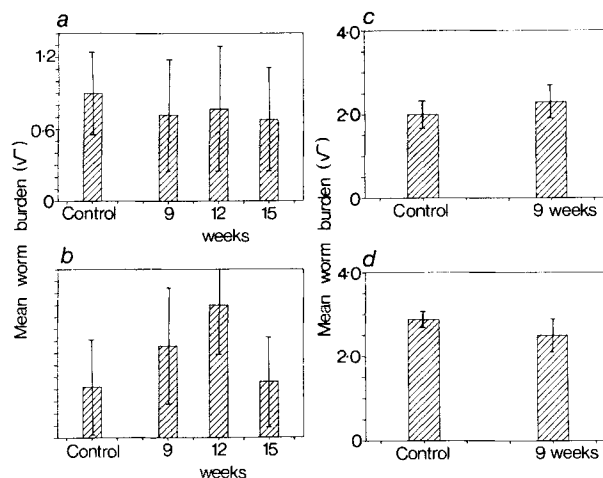


Fig. 2 Worm recovery following curative chemotherapy. Praziquantel was administered at a dose of 250 mg per kg divided into three applications given at 4-h intervals during a single day²¹.

a, The control was exposed to infections of 10 cercariae per group of mice per week for 6 weeks before autopsy. The experimental groups (10 mice per group) were exposed to an identical repeated infection for 9, 12 and 15 weeks before chemotherapy; autopsy was performed after 6 weeks of reinfection. One-way analysis of variance on transformed data (a square-root transformation based on the observed Poisson distributions of parasite numbers per mouse) showed no significant differences between treatments and control at the 5% level. Note, however, that the type II errors associated with this test (and those associated with the data presented in *b, c* and *d*) are large because of small sample sizes (10 mice per group). Studies involving much larger samples are required for future work. *b*, As for *a* except that worm recovery was performed 9 weeks after initial infection in the control and after 9 weeks reinfection following chemotherapy in the experimental groups. One-way analysis of variance on transformed data ($\sqrt{x}$) showed no significant differences between treatments and control at the 5% level. *c*, The control group was exposed to infections of 100 cercariae per group of mice per week for 6 weeks before autopsy. The experimental group was exposed to 9 weeks infection before treatment (100 cercariae per week) and to 6 weeks reinfection post treatment. Student's *t*-test ($\sqrt{x}$) showed no significant differences at the 5% level. *d*, As for *c* but with worm recovery performed 9 weeks after initial infection in the control group and after 9 weeks reinfection in the experimental group. No significant difference at the 5% level (*t*-test). *a–d*, Vertical lines show the 95% confidence limits. Note that the sample sizes of experimental animals (10 mice per group) were low and further studies are required to confirm these observations.

results (based on primary and single challenge experiments) are inconclusive¹⁰⁻¹²; residual protective immunity was apparent in some studies but not in others. Our results from continual exposure experiments tentatively suggest that the expulsion of adult worm burdens eliminates the 'memory' of previous infections.

Studies of parasite establishment and mortality in hosts continually exposed to helminth infection have many advantages over more conventional experimental designs when seeking information to help interpretation of observed epidemiological

trends in human communities. Where rates of parasite establishment depend on previous experience of infection, horizontal or longitudinal profiles of changes in parasite burden with age can vary from growth to a stable plateau in parasite load in areas of low transmission intensity to convex curves with marked declines in infection in older age classes in areas of high transmission intensity.

We thank the MRC (a research studentship to J.A.C.), the Rockefeller Foundation (R.M.A.) and the Edna McConnell Clark Foundation (R.M.A.) for financial support.

Received 3 December 1984; accepted 3 April 1985.

1. Wakelin, D. *Immunity to Parasites* (Arnold, London, 1984).
2. Butterworth, A. E., Taylor, D. W. & Veith, M. C. *Immun. Rev.* **61**, 5-39 (1982).
3. Cohen, S. & Warren, K. S. *Immunology of Parasitic Infections* (Blackwell Scientific, Oxford, 1982).
4. Mitchell, G. F. *Adv. Immun.* **28**, 451-467 (1980).
5. Anderson, R. M. & May, R. M. *Adv. Parasit.* **24**, 1-101 (1985).
6. Taylor, A. E. R. & Muller, R. (eds) *Br. Soc. Parasit. Symp.* **18** (Blackwell Scientific, Oxford, 1980).
7. Sturrock, R. F. *et al. Trans. R. Soc. trop. Med. Hyg.* **75**, 219-231 (1981).
8. Dean, D. A. *Expl Parasit.* **55**, 1-104 (1983).
9. Jenkins, D. C. & Phillips, R. F. *Parasitology* **62**, 457-465 (1970).
10. Jenkins, D. C. & Phillips, R. F. *Int. J. Parasit.* **2**, 105-111 (1972).

11. Sturrock, R. F., Cottrell, B. J. & Kimani, R. *Parasitology* **88**, 505-514 (1984).
12. Cheever, A. W., Kamel, J. A., Elwi, A. M., Musiman, J. E. & Danner, R. *Am. J. trop. Med. Hyg.* **26**, 702-716 (1977).
13. Scott, D., Senker, K. & England, E. C. *Bull. Wild Hlth Org.* **60**, 89-100 (1982).
14. Bradley, D. J. & McCullough, F. *Trans. R. Soc. trop. Med. Hyg.* **67**, 491-513 (1973).
15. Anderson, R. M. *Parasitology* **72**, 281-305 (1976).
16. Warren, K. S. *J. infect. Dis.* **127**, 595-609 (1973).
17. Dalton, P. R. & Pole, D. *Bull. Wild Hlth Org.* **563**, 417-424 (1978).
18. Andrade, Z. A. & Warren, K. S. *Trans. R. Soc. trop. Med. Hyg.* **58**, 53-57 (1964).
19. Anderson, R. M. & May, R. M. *Nature* **315**, 493-496 (1985).
20. Jordan, P. & Webbe, G. *Schistosomiasis: Epidemiology, Treatment and Control* (Heinemann, London, 1982).
21. Gonnert, R. & Andrews, P. Z. *Parasit.* **52**, 129-150 (1977).

Herd immunity to helminth infection and implications for parasite control

R. M. Anderson & R. M. May*

Department of Pure and Applied Biology,
Imperial College of Science and Technology,
Prince Consort Road, London SW7 2BB, UK

* Biology Department, University of Princeton, Princeton,
New Jersey 08540, USA

Despite much research on immunological responses to helminth parasites, knowledge of the dynamic interplay between levels of herd immunity in humans and the rates of exposure, establishment and mortality of parasites remains limited¹⁻⁷. We describe here a simple mathematical model for the population dynamics of helminth infections which mirrors the development of a degree of acquired immunity within populations which are genetically heterogeneous with respect to immunological responsiveness. We interpret observed patterns in the age-specific intensity of infection and attempt to understand the possible effects of control measures based on chemotherapy and vaccination. Mass chemotherapy can, in some circumstances, reduce the level of herd immunity such that average worm burdens in the adult age classes rise above their precontrol levels. When certain individuals or groups are predisposed to heavy infection^{5,8,9}, selective or targeted drug treatment can have significantly greater impact than mass or random application. Conversely, model predictions suggest that effective parasite control by vaccination (if and when vaccines become available) is difficult to achieve in communities that are genetically heterogeneous in their ability to mount protective responses to infection.

None of the existing epidemiological models of helminth transmission in human communities takes explicit account of acquired immunity, despite the growing volume of field¹⁰⁻¹⁶ and laboratory-based studies¹⁷⁻¹⁹ which attest to its significance as a determinant of parasite abundance. Infection with helminths and exposure to their diverse and complex antigens, generates varied and complex immune responses^{4,7,18,20}. Acquired immunity seems to be characterized generally by a marked dependence on the degree of antigenic stimulation (the accumulated experience of infection over time), an ability to provide only partial protection against reinfection, and a transient efficacy^{7,17-20}. Host responses seem to be most effective against the invading larval stages of helminth parasites and, as such, act to decrease the rate of parasite establishment within the host^{7,18}. In endemic areas, humans seem continually to acquire parasites throughout life, although the average intensity of infection is often less in older than in younger age groups²¹ (Fig. 1a,

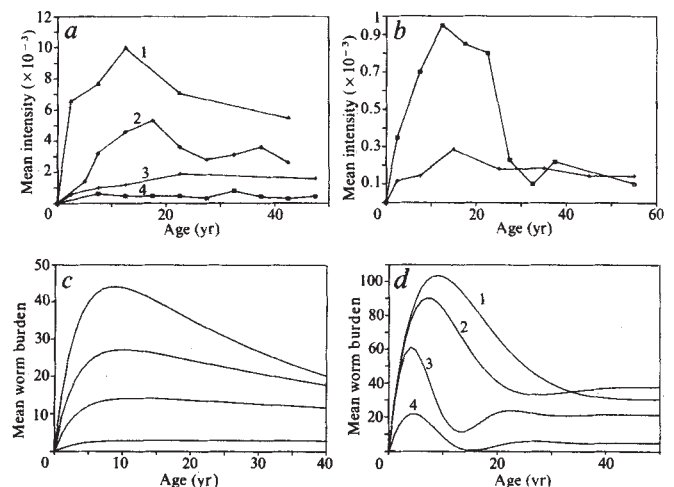


Fig. 1 Observed and predicted patterns of age-related changes in the average intensity of infection (horizontal cross-sectional profiles). **a**, Hookworm infection in communities with varying intensities of transmission. Data sources are: curves 1, 3, ref. 29; curve 2, ref. 30; curve 4, ref. 31. The indirect measure of hookworm burden (intensity) is mean eggs per g of faeces. **b**, Schistosome infection in areas of high³² and low³³ transmission intensity. Intensity measured by eggs per g of faeces (Kato technique). Note how in both **a** and **b** the intensity profiles become convex in areas of high transmission intensity. **c**, Equilibrium/age-intensity curves ($\partial M(t, a)/\partial t = 0$) generated by the model defined in equation (1) (with $\Delta = 0$) with the following parameters (population assumed to be homogeneous): $\epsilon = 0.0005$, $\sigma = 0.00001$, $\mu = 0.33$, $\Lambda = 17$ (top line); 10, 5 and 1 (bottom curve) (all rates are per yr). If the net infection rate in equation (1) < 0 (that is, ϵ large) acquired immunity is assumed to influence parasite establishment and survival. Note how the curves become convex in shape as the intensity of transmission increases (the value of Λ). **d**, Similar to **c** but with parameter values: curve 1, at age 10, $\Lambda = 10$, $\mu = 0.2$, $\epsilon = 0.009$, $\sigma = 0.05$; curve 2, at age 10, $\Lambda = 30$, $\mu = 0.2$, $\epsilon = 0.001$, $\sigma = 0.05$; curve 3, at age 10, $\Lambda = 30$, $\mu = 0.2$, $\epsilon = 0.004$, $\sigma = 0.1$; curve 4, $\Lambda = 30$, $\mu = 0.2$, $\epsilon = 0.005$, $\sigma = 0.02$ (all rates are per yr).

b). The degree to which intensity declines in the older people seems to be positively associated with the intensity of transmission (Fig. 1a, b). Much current debate centres on the relative roles of acquired immunity and age-specific changes in exposure to infection as determinants of such trends.

Current understanding of acquired immunity can be crudely captured by a model in which the rate of establishment of new parasites within individuals of a given age, a , is inversely correlated with some accumulated average experience of infection

Fig. 2 Heterogeneity in 'immunological responsiveness' and exposure to infection. Model predictions for a population consisting of two types of individuals who represent proportions f_1 and f_2 of the total population, where $f_1 + f_2 = 1$. where $M_1(t, a)$ is the mean worm load of type I individuals (f_1) and $M_2(t, a)$ is the mean worm load of type II individuals (f_2) at time, t , of age, a , at equilibrium ($\partial M_1(t, a)/\partial t = \partial M_2(t, a)/\partial t = 0$) the model takes the form

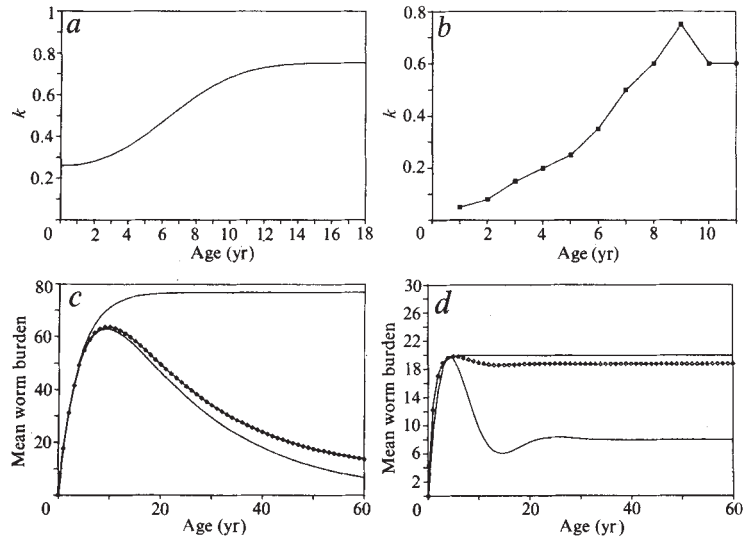
$$dM_1(a)/da = \Lambda_1 \left[1 - \varepsilon_1 \int_0^a M_1(a') e^{-\sigma_1(a-a')} da' \right] - \mu_1 M_1(a)$$

$$dM_2(a)/da = \Lambda_2 \left[1 - \varepsilon_2 \int_0^a M_2(a') e^{-\sigma_2(a-a')} da' \right] - \mu_2 M_2(a)$$

where the subscripts to the parameters denote differences in their values between type one and type two individuals. Here $\Lambda_i = R_{0i}(b + \sigma_i)M$ where $1/b$ is the life expectancy of the human host, R_{0i} is the basic reproductive rate of the parasite in hosts of class i and $\bar{M}$ is the overall average worm burden in the entire population (average over all age classes and types),

$$\bar{M} = \sum_{i=1}^2 f_i b \int_0^\infty M_i(a) e^{-ba} da$$

a, Predicted changes with age in the negative binomial parameter k (which varies inversely with the degree of parasite contagion within the host population)^{5,21}. Model parameters: $\Lambda_1 = 1$, $f_1 = 0.9$, $\mu_1 = \mu_2 = 1$, $\varepsilon_1 = \varepsilon_2 = 0.01$, $\sigma_1 = \sigma_2 = 0.1$, $\Lambda_2 = 20$, $f_2 = 0.1$; all are per yr. The differences in Λ reflect differing rates of exposure to infection (perhaps induced by behavioural differences between the two groups of individuals). **b**, Observed age-related changes in k for hookworm infection (*Necator americanus*) in a village community in India²⁴. **c**, Predicted age-related changes in the mean worm burdens in type I individuals ($M_1(a)$) (top solid line), type II individuals ($M_2(a)$) (bottom solid line) and the total population $M(a)$ (diamond symbols). This example mirrors a population in which the wormy people (fraction f_1) are few in number ($f_1 = 0.1$) and are characterized by a short immunological memory ($\sigma_1 = 1$). The dominant proportion of the population ($f_2 = 0.9$) have a long-lasting memory ($\sigma_2 = 0.00001$, that is, lifelong). Parameter values: $\Lambda_1 = \Lambda_2 = 20$, $\mu_1 = \mu_2 = 0.25$, $\varepsilon_1 = \varepsilon_2 = 0.05$; all per yr. **d**, Similar to **c** but the dominant proportion of the population are poor immunological responders to infection ($f_1 = 0.9$) with a low rate of acquisition of immunity ($\varepsilon_1 = 0.00001$) when compared with the small fraction of good responders ($\varepsilon_2 = 0.01$). Parameter values: $\Lambda_1 = 20$, $\Lambda_2 = 10$, $\mu_1 = 1$, $\mu_2 = 0.25$, $\sigma_1 = \sigma_2 = 0.1$; all per yr.



through time. Laboratory studies of mice repeatedly exposed to helminth infection provide some support for this view (see accompanying paper¹⁹). We make the simple assumption that the rate of parasite establishment declines linearly as the accumulated experiences of exposure to infection and worm load increase with age. 'Immunological memory' of these past experiences is assumed to last only for a finite period of time. These assumptions can be expressed as a single partial differential equation denoting changes in the mean worm burden, $M(t, a)$, with respect to time t and age a :

$$\frac{\partial M(t, a)}{\partial t} + \frac{\partial M(t, a)}{\partial a} = \Lambda(t) \left[1 - \Delta \int_0^a \Lambda(t') \exp[-\sigma_2(a-a')] da' \right] - \varepsilon \int_0^a M(t, a') \exp[-\sigma_1(a-a')] da' - \mu M(t, a) \quad (1)$$

Here $\Lambda(t)$ denotes that per capita rate of infection at time t ; it is a function of the average fecundity of mature worms, infective stage life expectancy (for indirectly transmitted infection this includes the life expectancy of short-lived vectors or intermediate hosts), host density and the probability of host contact with infective stages or infected vectors²¹. The parameters Δ and ε record the strength of acquired immunity as it acts to decrease parasite establishment resulting from either exposure to infection (Δ) or adult parasite burdens (ε); $1/\sigma_2$ and $1/\sigma_1$ reflect the average duration of the immunological memory of past exposure and past worm loads, respectively (if $\sigma_i = 0$, memory is life-long; if $\sigma_i = \infty$ memory is absent); and $1/\mu$ denotes the life expectancy of the adult parasites in man. Here, for simplicity, we set $\Delta = 0$ because past worm burdens partly reflect past exposure. However, equation (1) has the flexibility to allow for the somewhat different ways in which immunity is influenced by exposure to infection as well as by the burden of established worms.

The acquisition of immunity to parasite establishment is assumed to be the only major constraint on parasite population

growth (parasite fecundity and mortality are taken to be independent of parasite density in the present model; such effects are discussed elsewhere^{5,9,21,22}). Equation (1) assumes that the human population is homogeneous in its immunological response to parasite invasion (that is, ε is a constant). This assumption can be relaxed by dividing the population into groups ($f_1 \dots f_n$) in which ε , μ , Λ and σ adopt different values (Λ_i , ε_i , μ_i and σ_i). At equilibrium ($\partial M(t, a)/\partial t = 0$), the mean worm burden at age a , $M(a)$, is simply (with $\Delta = 0$ and λ real¹⁹),

$$M(a) = A + (\bar{\Lambda}/\lambda) \left[\frac{(p_1 + \sigma)}{p_1} e^{p_1 a} - \frac{(p_2 + \sigma)}{p_2} e^{p_2 a} \right] \quad (2)$$

where

$$\begin{aligned} A &= \bar{\Lambda} \sigma / (\sigma \mu + \varepsilon \bar{\Lambda}); & p_1 &= (\lambda - \mu - \sigma)/2; \\ p_2 &= -(\lambda + \mu + \sigma)/2; & \lambda &= ([\mu - \sigma]^2 - 4\varepsilon \bar{\Lambda})^{1/2}; \\ \bar{\Lambda} &= (\sigma + b)(R_0 - 1)(b + \mu)/\varepsilon; \end{aligned}$$

and $1/b$ = human life expectancy (assuming a type II survival curve). The quantity R_0 denotes the basic reproductive rate of the parasite^{5,9,21,22} ($R_0 > 1$ for parasite persistence). Stochastic models for the distribution of worms among hosts in the equilibrium state reveal that the probability distribution of worm numbers per host is Poisson if ε , σ and μ are constant within the human community (the homogeneous case) but overdispersed if R_0 , σ or μ vary within different segments of the population (heterogeneity in immunological responsiveness or exposure to infection). Although crude parameter estimates are available for Λ (from studies of reinfection following chemotherapeutic treatment) and μ , little is known about the possible values of ε and σ for the major helminth infections of man. However, the equilibrium/age-intensity curves generated by the model for various plausible values of ε and σ are similar to observed patterns (Fig. 1a-d). The model can mimic a range of patterns including: monotonic increase in $M(a)$ to a stable plateau; $M(a)$ peaking within childhood, teenage and early adult age groups, and then declining; and damped oscillations in intensity (Fig. 1a-d). For a fixed severity of the acquired response (ε), the worm burden curve tends to be convex if memory is long,

and/or transmission is intense, and/or worm life expectancy is long. The age at which a maximum worm load is attained depends primarily on the life expectancy of the worm ($1/\mu$) and the severity of acquired immunological response (ε). Certain of these predictions are in qualitative agreement with observed trends: a decline in $M(a)$ in older age groups is more apparent in areas of intense transmission (Fig. 1a, b). The damped oscillations in mean worm burden that can be produced within individuals under appropriate circumstances derive from essentially the same mechanisms that can produce oscillations in the incidence of microparasitic infections, such as measles, within populations²³.

Observed distributions of parasite number per person are invariably highly aggregated, with most people harbouring few worms and a few people harbouring most parasites (often 70% of the worms are harboured by <10% of the community)^{5,8,9,21,22}. For certain helminths, these 'wormy' people seem to be predisposed to this state by as yet undetermined genetic, behavioural or social factors^{8,9}. Under the assumption that predisposition to heavy (or light) infection is primarily under genetic²⁴⁻²⁸ (via the immune system) or behavioural control, our model can be extended to encompass variability in the way different groups of individuals within a community either respond to parasitic invasion (different values of the parameters ε , σ and μ) or are exposed to infection (different values of Λ).

The range of patterns generated by such 'heterogeneous models' is bewilderingly large. Here, we simply focus on a model with two types of people; namely, high responders (resistive individuals who harbour low burdens) and low responders ('wormy' people). Figure 2 indicates a range of possible patterns for the change in intensity of infection with age. Heterogeneity between the two groups generates aggregation of parasite numbers within the total population, as measured inversely by the negative binomial dispersion parameter, k (Fig. 2a). Heterogeneity in exposure to infection (Λ), via its interaction with the rate of acquisition of immunity, can generate age-related changes in the degree of parasite contagion within age classes; these patterns are similar to observed trends (Fig. 2a, b). Oscillatory changes in worm load with age in high responders can be masked within the overall average burden if low responders predominate and, conversely, the severity of infection within 'wormy' people may be hidden by the overall average if high responders predominate (Fig. 2c, d). Therefore, average measures may hide interesting age-related trends in individuals predisposed to heavy or light infection. Such trends, if observable (by longitudinal studies of individual patients), could provide valuable insights into the nature of herd immunity to helminth infection.

The dynamics of naturally acquired immunity have important implications for control programmes. Repeated mass chemotherapy at a level less than that required to eradicate the parasites can significantly reduce the level of herd immunity and can raise average worm burdens in the older age classes above the levels pertaining before control (Fig. 3a). This trend will be most apparent if strong, long-lasting immunity is acquired (where σ is small and ε large) in areas where transmission is intense (especially relevant to schistosome and filarial infections). This danger can be avoided by targeting drug treatment to those predisposed to heavy infection (see, for example, Fig. 3b). Although helpful for community-based chemotherapy, predisposition to heavy infection as a consequence of poor immunological competence may present serious problems for parasite control by vaccination (if and when vaccines become available). A vaccine must be able to convert low responders (wormy people) into high responders if vaccination on a community scale is to be successful.

The model is a crude approximation to the true complexity of acquired immunity to helminths. Although it generates various epidemiological patterns that are similar to those actually observed, predictions must remain tentative; the main purpose of theoretical analysis is to highlight some of the poss-

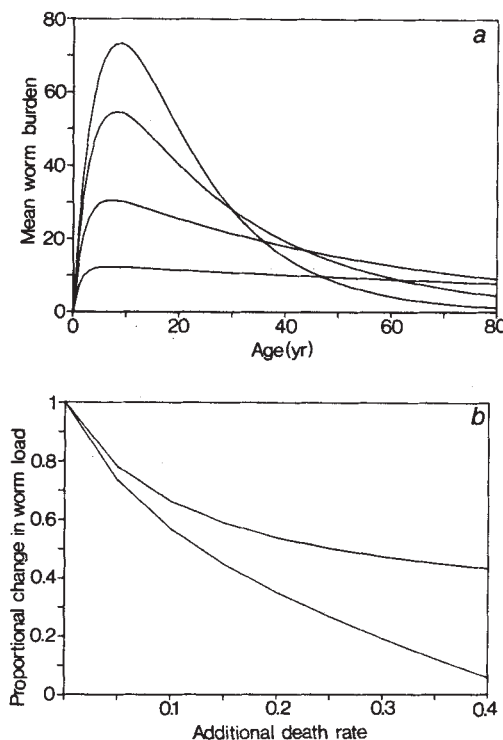


Fig. 3 Impact of chemotherapy on the intensity of infection. **a**, Homogeneous population (with respect to immunological responsiveness and exposure to infection) in which drug treatment is applied randomly (independent of parasite load) and repeatedly. The four curves denote the equilibrium/age-intensity profile for the precontrol situation (curve 1, at age 10) and three levels of drug application which increase the overall death rate of the parasite within the population from $\mu = 0.25$ (precontrol) to $\mu = 0.33$ (curve 2, at age 10), $\mu = 0.5$ (curve 3, at age 10) and $\mu = 0.75$ (curve 4, at age 10). Parameter values: $\Lambda(\mu) = (\sigma + b)(b + \mu)[R_0 - 1]$, $b = 1/70$, μ = variable, $\varepsilon = 0.0005$, $\sigma = 0.00001$, $R_0 = 1.44$ (when $\mu = 0.75$), $R_0 = 2.13$ (when $\mu = 0.5$), $R_0 = 3.19$ (when $\mu = 0.33$) and $R_0 = 4.16$ (when $\mu = 0.25$). **b**, Proportional change at equilibrium (over that pertaining before control) in the overall population worm load, $\bar{M}$ (over all age classes and all types of individual) in a population subject to varying degrees of treatment targeted at a 'wormy' fraction ($f_1 = 0.1$). The non-wormy people ($f_2 = 0.9$) are untreated. The value of $\bar{M}$ was calculated from the equation

$$1 = \frac{f_1 R_{01}(b + \sigma_1)}{[(b + \sigma_1) + \varepsilon_1 R_{01} \bar{M}]} + \frac{f_2 R_{02}(b + \sigma_2)}{[(b + \sigma_2) + \varepsilon_2 R_{02} \bar{M}]}$$

where $b = 1/70$, $\varepsilon_1 = 0.00001$, $\varepsilon_2 = 0.0075$, $\sigma_1 = \sigma_2 = 0.1$, $\mu_1 = 0.2$, $\mu_2 = 0.6$, $R_{01} = 4.67$, $R_{02} = 1.63$. Horizontal axis denotes the extra death rate (above the precontrol value of μ_1 and μ_2) resulting from mass chemotherapy. Curve 1 denotes the result of targeting in which only a fraction ($f_1 = 0.1$) are treated (the wormy people with a low ε_1 value, a low parasite death rate (μ_1) and a high parasite basic reproductive rate (R_{01}). Curve 2 denotes mass treatment of the entire population. Note that for an additional death rate of 0.3 per yr (on μ_1 in the case of targeted treatment, and on μ_1 and μ_2 in the case of mass application) the selective treatment of 10% of the population gives a reduction in average worm load ($\bar{M}$) to ~50% (curve 1). By treating the entire population, only an additional 30% change is attained (curve 2).

ible population-level implications of acquired immunity to helminth infection. In particular, the model indicates how valuable insights may be gained from longitudinal epidemiological studies of fluctuations in parasite load within individual patients¹⁰⁻¹⁶ and from experimental studies of long-term continual infection in laboratory animals, described in the accompanying paper¹⁹.

We thank the Rockefeller Foundation, the Edna McConnell Clark Foundation (R.M.A.) and the NSF (R.M.M.) for financial support. We also thank R. M. Maizels, J. A. Crombie and G. A. Schad for useful discussions.

Received 1 October 1984; accepted 25 April 1985.

- Peters, W. in *The Relevance of Parasitology to Human Welfare Today* (eds Taylor, A. E. & Muller, R.) 1-17 (Blackwell Scientific, Oxford, 1978).
- Walsh, J. A. in *Tropical and Geographic Medicine* (eds Warren, K. S. & Mahmoud, A. A. F.) 1073-83 (McGraw-Hill, New York, 1984).
- Lloyd, S. *Parasitology* **83**, 225-242 (1981).
- Wakelin, D. *Crit. Rev. trop. Med.* **2** (in the press).
- Anderson, R. M. & May, R. M. *Nature* **287**, 557-563 (1982).
- Mahmoud, A. A. F., Siongok, T. K. A. & Ouma, J. *Lancet* **i**, 849-851 (1983).
- Wakelin, D. *Immunity to Parasites* (Arnold, London, 1984).
- Schad, G. A. & Anderson, R. M. *Science* (in the press).
- Anderson, R. M. & Medley, G. F. *Parasitology* (in the press).
- Butterworth, A. E. *et al. Trans. R. Soc. trop. Med. Hyg.* **78**, 108-123 (1984).
- Butterworth, A. E. *et al. Trans. R. Soc. trop. Med. Hyg.* (in the press).
- Sturrock, R. F. *et al. Trans. R. Soc. trop. Med. Hyg.* **77**, 363-371 (1983).
- Sturrock, R. F., Cottrell, B. J. & Kimani, R. *Parasitology* **88**, 505-514 (1984).
- Wilkins, H. A., Goll, P. H., Marshall, T. F. & Moore, P. J. *Trans. R. Soc. trop. Med. Hyg.* **78**, 216-221 (1984).
- Goll, P. H., Wilkins, H. A. & Marshall, T. F. *Trans. R. Soc. trop. Med. Hyg.* **78**, 222-226 (1984).
- Wilkins, H. A., Goll, P. H., Marshall, T. F. & Moore, P. J. *Trans. R. Soc. trop. Med. Hyg.* **78**, 227-232 (1984).
- Mitchell, G. F. *Adv. Immun.* **28**, 451-511 (1979).
- Dean, D. A. *Expl Parasit.* **55**, 1-104 (1983).
- Crombie, J. A. & Anderson, R. M. *Nature* **315**, 491-493 (1985).
- Miller, T. A. *Adv. Parasit.* **17**, 315-383 (1979).
- Anderson, R. M. & May, R. M. *Adv. Parasit.* **24**, 1-101 (1985).
- Anderson, R. M. in *Population Dynamics of Infectious Diseases: Theory and Application* (ed. Anderson, R. M.) Ch. 3 (Chapman & Hall, London, 1982).
- Anderson, R. M. & May, R. M. *Science* **215**, 1053-1060 (1982).
- Skamene, E., Kongschaun, P. A. L. & Landy, M. (eds) *Genetic Control of Natural Resistance to Infection and Malignancy* (Academic, New York, 1980).
- Wakelin, D. in *Animal Models in Parasitology* (ed. Owen, D. G.) 53-63 (Macmillan, London, 1982).
- Bickle, Q., Long, E., James, E., Daenhoff, M. & Festing, M. *Expl Parasit.* **50**, 222-232 (1980).
- Abdel-Salem, E., Ishaac, S. & Mahmoud, A. A. F. *J. Immun.* **123**, 1829-32 (1979).
- Bodmer, W. F. *J. R. Coll. Phys.* **14**, 43-50 (1980).
- Sadun, E. H. & Vajrashira, S. *Am. J. trop. Med. Hyg.* **2**, 286-297 (1953).
- Carr, H. P. *Am. J. Hyg.* **6**, 42-61 (1926).
- Pesigan, T. P. *et al. Bull. Wild Hlth Org.* **18**, 345-455 (1958).
- Siongok, T. K. A. *et al. Am. J. trop. Med. Hyg.* **25**, 273-284 (1976).
- Abdel-Wahab, M. F. *et al. Am. J. trop. Med. Hyg.* **29**, 868-874 (1980).
- Anderson, R. M. *Lect. Notes Biomath.* **39**, 278-322 (1980).

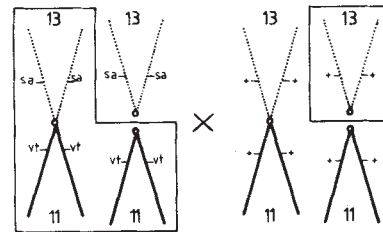


Fig. 1 Production and detection of disomy 11 and 13. Heterozygotes for the robertsonian translocation, Rb(11.13)4Bnr, exhibit high levels of non-disjunction of chromosomes 11 and 13. When intercrossed, normal disjunction in both parents results in the production of chromosomally balanced young; with non-disjunction in one parent, monosomy and trisomy for chromosomes 11 and/or 13 result; but when non-disjunction occurs in both parents and complementary products (as illustrated) unite, chromosomally balanced young disomic for chromosomes 11 (as shown) or 13 are produced. These are most readily detected when one parent is homozygous for a gene marker; marked young are then either maternal or paternal disomics according to which parent carries the marker. Disomy 11 is detected in this way with a frequency of about 1-2%¹⁶ using the gene marker vestigial tail (*vt*). It has also been found with a dominant marker, Wavy coat (*Re^{wc}*). Marked maternal disomics are smaller than their sibs, marked paternal disomics are larger. Most unmarked young are of normal chromosome origin but unmarked disomic young are also found; in crosses producing large paternal disomy 11 progeny, small maternal disomy 11 young may also occur (*x* and *y*, Fig. 2) and their genotype can be confirmed by breeding tests. The reverse is also true. The size phenomenon is not therefore dependent on the marker genes present. In the same crosses, satin (*sa*) was the marker for detecting disomy 13.

Differential activity of maternally and paternally derived chromosome regions in mice

B. M. Cattanch & M. Kirk

MRC Radiobiology Unit, Harwell, Didcot, Oxon OX11 0RD, UK

Although both parental sexes contribute equivalent genetic information to the zygote, in mammals this information is not necessarily functionally equivalent. Diploid parthenotes possessing two maternal genomes are generally inviable¹, embryos possessing two paternal genomes in man may form hydatidiform moles², and nuclear transplantation experiments in mice have shown that both parental genomes are necessary for complete embryogenesis³⁻⁶. Not all of the genome is involved in these parental effects, however, because zygotes with maternal or paternal disomy for chromosomes 1, 4, 5, 9, 13, 14 and 15 of the mouse survive normally^{7,8}. On the other hand, only the maternal X chromosome is active in mouse extraembryonic membranes⁹, maternal disomy 6 is lethal⁷, while non-complementation of maternal duplication/paternal deficiency or its reciprocal for regions of chromosome 2, 8 and 17 has been recognized¹⁰⁻¹². We report that animals with maternal duplication/paternal deficiency and its reciprocal for each of two particular chromosome regions show anomalous phenotypes which depart from normal in opposite directions, suggesting a differential functioning of gene loci within these regions. A further example of non-complementation lethality is also reported.

The first set of opposite phenotypes was found in mice disomic for chromosome 11. These were produced, together with animals with disomy 13, by intercrossing heterozygotes for the robertsonian translocation, Rb(11.13)4Bnr, which typically exhibit high levels of non-disjunction of chromosomes 11 and 13 (ref. 13) (Fig. 1). Disomy 13 mice appeared normal in all respects, like other disomics previously produced^{7,8}, but disomy 11 mice were found to differ markedly in size from their sibs; sig-

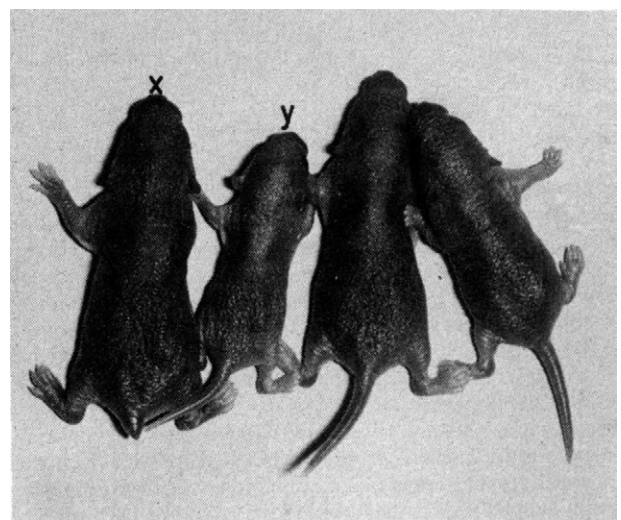


Fig. 2 Maternal and paternal disomy 11 in a single litter. Litter from an unmarked Rb4Bnr female heterozygote sired by a *vt* marked Rb4Bnr male heterozygote. The *vt* marked paternal disomic (*x*) is large, whereas one marked sib is clearly small (*y*). Breeding tests on such small animals show that most are not heterozygous for *vt*, but rather homozygous wild type and therefore maternal disomics. The reverse situation also occurs. Marked and unmarked disomics occur with near-equal frequencies, although seldom, as here, within single litters.

nificantly, when their chromosomes 11 were of maternal origin they were consistently smaller than their littermates, and when their chromosomes 11 were of paternal origin they were consistently larger. This was true for both sexes. The phenomenon was most evident within a few hours of birth (Fig. 2), although the size differences present at birth (ratio of weights, 0.606 ± 0.015 g, $n = 40$, and 1.408 ± 0.036 g, $n = 36$) usually persist to adulthood (small ♀♀, 0.576 ± 0.036 g, $n = 11$; small ♂♂, $0.627 \pm$

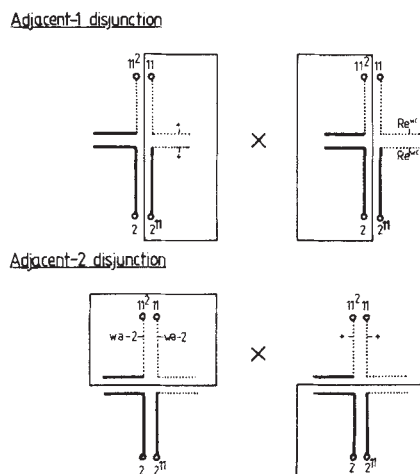


Fig. 3 Intercrosses of T(2;11)30H heterozygotes. Alternate segregation in both parents leads to the production of chromosomally balanced young; adjacent-1 or adjacent-2 disjunction in one parent leads to the production of young with unbalanced, duplication/deficiency chromosomal constitutions. However, when adjacent-1 disjunction occurs in both parents and complementary products (as illustrated) unite, chromosomally balanced young appear which are either maternal duplication/paternal deficiency or maternal deficiency/paternal duplication for the distal region of chromosome 11 (and vice versa for chromosome 2). These can readily be detected when one parent is homozygous for a gene marking this chromosome region, Re^{wc} , in the present experiments. Each of the two duplication/deficiency types occurs with a frequency of about 11–14% ($n = 514$). When adjacent-2 disjunction occurs in both parents and complementary products (as illustrated) unite, chromosomally balanced young which are either maternal duplication/paternal deficiency or maternal deficiency/paternal duplication for the proximal region of chromosome 11 (and vice versa for chromosome 2) should be produced. These should be detectable with a proximal 11 chromosome marker, *waved-2* (*wa-2*), in the present experiments. When the female parent carried the marker, *wa-2* progeny representing maternal duplication/paternal deficiency for proximal 11 were detected with a frequency of ~1.3% (2/153 classified for *wa-2* at 10 days of age); all were small like maternal disomy 11. No large unmarked young which could have represented maternal deficiency/paternal duplication for proximal 11 were detected, however, and, consistent with this, no *wa-2* progeny were recovered (0/248) from intercrosses in which the male parent carried the marker. The genotype may therefore be lethal (see text). However, small, unmarked young which proved to be maternal duplication/paternal deficiency types by breeding tests were recovered from the latter cross (3/248).

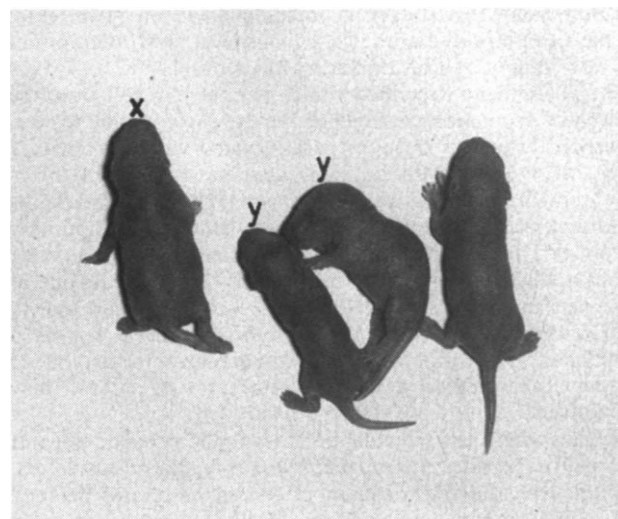


Fig. 4 Anomalous young produced from T(2;11)30H intercrosses. x, Phenotype characterizing maternal deficiency/paternal duplication for the distal region of chromosome 2; y, phenotype characterizing maternal duplication/paternal deficiency for this region (see text).

As paternal disomy 11 is viable, the lethality may be reasonably attributed to the chromosome 2 constitution (maternal duplication/paternal deficiency for proximal 2) and represent a further example of the non-complementation phenomenon described by Searle and Beechey¹⁰.

The second set of opposite phenotypes was derived from the T30H crosses. Although both maternal duplication/paternal deficiency for distal 11 and its reciprocal were detected, both genotypes proved to be postnatal lethals. Again, because both maternal and paternal disomy 11 are viable, the lethality may be attributed to the chromosome 2 constitutions, namely, maternal deficiency/paternal duplication for distal 2 and its reciprocal. A lethality for maternal duplication/paternal deficiency for this region of chromosome 2 has been reported previously¹⁰, but a characteristic phenotype can now be described. Such mice have long flat-sided bodies with arched backs (Fig. 4). More strikingly, they become almost totally inactive within a few hours of birth and are killed by their mothers. Very few survive more than 24 h. A remarkable finding was that the reciprocal type, maternal deficiency/paternal duplication for distal 2, shows an effectively opposite phenotype. These mice have short square bodies with broad flat backs (Fig. 4) and are notably hyperkinetic. They usually survive for several days but fail to grow normally, and their behaviour becomes increasingly extreme, tremor and balance defects appearing frequently.

Because the anomalous phenotypes associated with maternal and paternal disomy 11 and corresponding types of duplication/deficiency for distal 2 can each occur among the progeny of single crosses, maternal cytoplasmic effects cannot be responsible for the abnormalities. Rather, a differential functioning of maternal and paternal chromosomes or chromosome regions is indicated, which suggests the existence of a form of chromosome imprinting^{14,15} that affects gene activity. The contrasting phenotypes (small and large; hypokinetic and hyperkinetic) depart from the normal in opposite directions and this suggests excess versus shortage of gene activity. This could mean either single or earlier activity of the paternal chromosome regions in the cases cited, and inactivity or later activity of the corresponding maternal regions. Paternal duplication would then give excess/earlier activity and the reciprocal subnormal/late activity. The X-inactivation process¹⁶ provides precedence for the activity of only one parental chromosome or chromosome

0.045 g, $n = 19$; large ♀♀, 1.328 ± 0.020 g, $n = 15$; but for large ♂♂, 1.234 ± 0.036 g, $n = 14$).

The large and small disomics were proportionately normal in all respects, they exhibited no viability problems and both sexes were fertile. The size differences did not persist into the next generation; the progeny of both large and small disomics of each sex and those of control Rb4Bnr heterozygotes were all of similar weight at birth and at 3 weeks of age. The absence of any indication of a maternal effect on progeny size may be attributed to the small litter sizes in all crosses, resulting from the high rates of translocation-associated non-disjunction.

Intercrosses between mice heterozygous for the reciprocal translocation, T(2;11)30H, provided evidence that only the proximal region of chromosome 11 (proximal 11) was associated with the size phenomenon (Fig. 3). Neither maternal duplication/paternal deficiency for distal 11, nor the reciprocal, resulted in mice which were of abnormal size at birth. However, maternal duplication/paternal deficiency for proximal 11 resulted in mice which were small in size both at birth and subsequently, resembling maternal disomy 11 in all respects. No young representing maternal deficiency/paternal duplication for proximal 11 have been found and it seems probable that this genotype is lethal.

region¹⁷; precedence for an earlier activity of one chromosome or chromosome region over its homologue has not been found, but has been reported at the gene locus level in diverse species and inter-family hybrids, including the mouse¹⁸⁻²³.

The phenomena described could, of course, result from the differential regulation of single loci in the chromosome regions concerned. The size phenomena associated with proximal 11 might involve a growth hormone gene such as Ames dwarf (*df*)²⁴, which lies in this chromosome region. The *df* mutant is of normal size at birth despite a severe deficiency of pituitary hormones²⁵, but the homologous chromosome region in man²⁶ carries a cluster of growth hormone loci²⁷, and at least one of these acts in the placenta. Hormonal control of fetal growth mediated through the placenta is therefore suggested. For distal 2, the behavioural problems of the anomalous mice may mean they have neurological defects; therefore, the *wst* locus²⁸ may be involved.

Single or preferential activation of regulator loci or of receptor loci, to give the latter different affinities for regulator substances, may underlie the phenomenon presented here, and detailed studies of the anomalous mice may provide some answers. An interaction between loci in different chromosome regions is a further possibility and studies with translocations having other breakpoints in chromosomes 2 and 11 could be informative here. Whatever the mechanism operating, it may well be the basis of the non-complementation lethality phenomenon¹⁰ and the lethality manifested by parthenogenic and androgenic embryos¹⁻⁶.

Since submission of this paper, it has been reported that an 180-base-pair DNA sequence thought to control pattern formation in early development, a homoeo box, is located on human chromosome 17 and mouse chromosome 11 (refs 29, 30) and therefore in the chromosome region responsible for the size phenomena described here. Another homoeo box has been located on chromosome 6 (ref. 31), which exhibits non-complementation lethality⁷. The occurrence of homoeo boxes in chromosomes which show differential activity according to parental origin could be fortuitous, but the association suggests a functional relationship. If true, homoeo boxes should also be found on chromosomes which show the parental origin phenomenon (2, 8 and 17), but perhaps not on those which do not (1, 4, 5, 9, 13, 14 and 15).

Received 21 January; accepted 25 March 1985.

- Markert, C. L. *J. Hered.* **73**, 390-397 (1982).
- Kajiji, T. & Ohama, K. *Nature* **268**, 633-634 (1977).
- Surani, M. A. H., Barton, S. C. & Norris, M. L. *Nature* **308**, 548-550 (1984).
- Barton, S. C., Surani, M. A. H. & Norris, M. L. *Nature* **311**, 374-376 (1984).
- McGrath, J. & Solter, D. *J. exp. Zool.* **228**, 355-362 (1983); *Nature* **308**, 550-551 (1984); *Cell* **37**, 179-183 (1984).
- Mann, J. R. & Lovell-Badge, R. H. *Nature* **310**, 66-67 (1984).
- Lyon, M. F. in *Radiation-Induced Damage in Man* (ed. Ishihara, T.) 327-346 (Liss, New York, 1983).
- Lyon, M. F., Ward, H. C. & Simpson, G. M. *Genet. Res.* **26**, 283-295 (1976).
- Tagaki, N. in *Preferential Inactivation of the Paternally Derived X Chromosome in Mice* (ed. Russell, L. B.) 341-360 (Plenum, New York, 1978).
- Searle, A. G. & Beechey, C. V. *Cytogenet. Cell Genet.* **20**, 282-303 (1978).
- Lyon, M. F. & Glenister, P. H. *Genet. Res.* **29**, 83-92 (1977).
- Johnston, D. R. *Genet. Res.* **24**, 207-213 (1975).
- Cattanach, B. M. & Moseley, H. *Cytogenet. Cell Genet.* **12**, 264-287 (1973).
- Crouse, H. V. *Genetics* **45**, 1429-1443 (1960).
- Brown, S. & Chandra, H. S. in *Cell Biology Vol. 1* (eds Golstein, L. & Prescott, D. M.) 109-189 (Academic, New York, 1977).
- Lyon, M. F. *Nature* **190**, 372-373 (1961).
- Burgoyne, P. S. in *Basic Reproductive Medicine Vol. 1* (eds Hamilton, D. & Naftolin, F.) 1-33 (MIT Press, 1981).
- Ohno, S. *Wistar Inst. Symp. Monogr.* **9**, 137-150 (1969).
- Schmidtke, J., Kuhl, P. & Engel, W. *Nature* **260**, 319-320 (1976).
- Klose, J. & Wolf, U. *Biochem. Genet.* **4**, 87-92 (1970).
- Whitt, G. S., Cho, P. L. & Childers, W. F. *J. exp. Zool.* **179**, 271-282 (1972).
- Whitt, G. S., Childers, W. F. & Cho, P. L. *J. Hered.* **64**, 55-61 (1973).
- Krietsch, W. K. G. *et al. Differentiation* **23**, 141-144 (1982).
- Green, M. C. in *Genetic Variants and Strains of the Laboratory Mouse* (ed. Green, M. C.) 8-278 (Gustav Fischer, Stuttgart, 1981).
- Bartke, A. *Gen. comp. Endocr.* **5**, 418-426 (1965).
- Buckle, V. J. *et al. Clin. Genet.* **26**, 1-11 (1984).
- Owerbach, D., Rutter, W. J., Martial, J. A., Baxter, J. D. & Shows, T. B. *Science* **209**, 289-292 (1980).
- Schultz, L. D., Sweet, H. O., Davison, M. T. & Coman, D. R. *Nature* **297**, 402-404 (1982).
- Joyner, A. L. *et al. Nature* **314**, 173-175 (1985).
- Rabin, M. *et al. Nature* **314**, 175-178 (1985).
- McGinnies, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. *Cell* **38**, 675-680 (1984).

Substance P raises neuronal membrane excitability by reducing inward rectification

P. R. Stanfield*, Yasuko Nakajima†
& Kazuhiko Yamaguchi

Department of Biological Sciences, Purdue University,
West Lafayette, Indiana 47907, USA

Much interest has recently centred on the properties of peptides that modulate the excitability of nerve cells. Such compounds include the undecapeptide substance P, which is particularly well established as an excitatory neurotransmitter^{1,2}, and we examine here its effects on magnocellular cholinergic neurones taken from the medial and ventral aspects of the globus pallidus of newborn rats and grown in dissociated culture³. These neurones have previously been shown to respond to substance P³ and are analogous to the nucleus basalis of Meynert in man^{4,5}, which gives a diffuse projection to the cerebral cortex and whose degeneration is the likely cause of Alzheimer's disease⁵. Substance P depolarizes these cultured neurones by reducing an inwardly rectifying potassium conductance; this conductance has been found in several neuronal types⁶⁻⁹ and has similar properties to those of certain other cells¹⁰⁻¹³. As discussed below, modulation of inward (or anomalous) rectification by substance P implies a self-reinforcing element to the depolarization caused by the peptide.

Neurones grown from 10 days to 3 weeks in culture were examined under whole cell recording^{14,15} (for details, see Fig. 1 legend). When substance P (3 μ M) was pressure ejected onto neurones examined in constant current conditions, a depolarization was produced (Fig. 1a) which lasted for 15 s to ≥ 5 min. Its amplitude was rarely as large as shown in Fig. 1a, but generally exceeded 1-2 mV. As described previously, this depolarization was associated with a fall in membrane conductance (for example, refs 16, 17), but in the cells examined in the present study, the fall in conductance was seen more clearly if the membrane was hyperpolarized (Figs 1b, 2a), implying that inward rectification⁶⁻¹³ is the conductance affected.

In establishing that inward rectification is modulated, we have sought to satisfy three criteria; first, that a K⁺ conductance with an inwardly rectifying current-voltage relation is reduced; second, that the conductance depends on external K⁺ concentration as well as on voltage^{11-13,18}; and third, that the conductance mechanism is nearly impermeable to rubidium (refs 10, 12, 19), which should therefore abolish the response to substance P.

The nature of the response to substance P is illustrated in the current-voltage relations of Fig. 2b. The substance P-sensitive current reversed its polarity close to E_K , the equilibrium potential for K⁺, reversal changing from -98.0 ± 1.9 mV ($n = 6$) to -85.4 ± 1.3 mV ($n = 14$) to -57.9 ± 1.6 mV ($n = 7$) as external K⁺ was changed from 2.5 to 5.0 to 12.5 mM (internal K⁺, 128 mM).

Although it remains possible that conductances other than inward rectification are affected, neither the transient A-current²⁰ (Fig. 2a) nor the delayed K⁺-current seem to be involved in the response of these cells. As Ca²⁺-activated K⁺ currents also become activated under depolarization^{21,22}, rather than under hyperpolarization as here, their involvement is unlikely. In addition, in our experiments internal Ca²⁺ was buffered to 10^{-7} M. The M-current, which is known to be involved in the response to substance P in certain cells²³, was not seen in the conditions of our experiments.

Our experimental external solutions contained tetrodotoxin (TTX, 1 μ M) to block Na⁺ currents, and blockage of Ca²⁺ currents with cadmium (0.5 mM, two cells) did not alter the response. Hence, the shape of the substance P-sensitive current-

* Present address: Department of Physiology, University of Leicester, Leicester LE1 7RH, UK.
† To whom correspondence should be addressed.

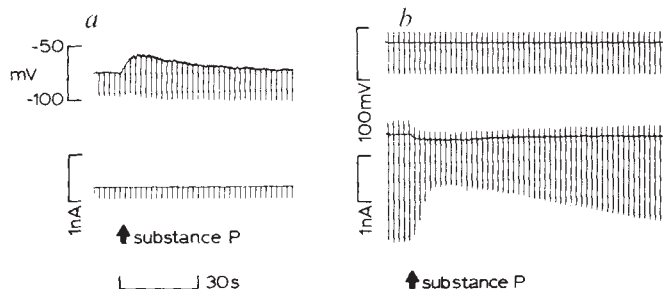


Fig. 1 Depolarization (a) and change in membrane current (b), measured under voltage clamp, resulting from application of 3 μ M substance P to a neurone cultured from the magnocellular basal forebrain of a newborn rat. The cell had been grown for 11 days in modified Eagle's medium with Earle's salt at 37 °C in a 10% CO₂ atmosphere. The methods of isolation, dissociation and culture have been briefly described previously³; a detailed account of the methods will be published elsewhere. The initial resting potential of the cell was -65 mV; soma diameter, 28 μ m; temperature, 31.2 °C. In a, a steady current of -75 pA was applied, which took the resting potential to approximately -75 mV; current pulses were -240 pA in amplitude and 100 ms long. In b, the holding potential was -76 mV. The depolarizing pulse was 20 mV in amplitude and the hyperpolarization, 60 mV. Both were 100 ms long. E_K , measured from the substance P response, was -89 mV. Whole cell recording was used^{14,15}. Recording pipettes were made from Blue Tip microhaematocrit capillary, and their tips were fire-polished but uncoated¹⁴. The pipette used in the experiment shown here had a resistance of 3.0 M Ω . The pipette (intracellular) solution contained (mM)¹⁵: K-aspartate, 120; NaCl, 40; MgCl₂, 3; CaCl₂, 0.25 and EGTA, 0.5 (calculated ionized Ca²⁺, 10⁻⁷ M); HEPES/KOH buffer, 5; Na₂ATP (vanadate-free), 2. The pH of this solution was adjusted to 7.2 with KOH. External solution (mM): KCl, 5; NaCl, 145.5; CaCl₂, 2.4; MgCl₂, 1.3; (+)glucose, 11; HEPES/NaOH buffer (pH 7.4), 5. Internal and external solutions were designed to have the same osmolality (~310 mosmol). The recording amplifier was a List electronic type EPC 7 patch clamp, and currents were filtered at 3 kHz, using the filter of this amplifier. Records of current, voltage and timing of substance P application were stored on a Hewlett-Packard 3520B tape recorder for later analysis. Voltage commands were provided from the gated output of a Digitimer type D4030, divided appropriately. Application of substance P was made by pressure ejection. A pipette, tip diameter 5 μ m, was filled with external solution containing substance P at 3 μ M. Timed pressure pulses (~0.2 p.s.i. and 1.63 s long) ejected this solution onto the cell. To avoid excessive depolarization of neurites by substance P, this pipette was placed close to the soma of the neurone. In this experiment the distance was 12.8 μ m from the edge of the soma, and single 'puffs' of substance P were applied at the times indicated by arrows.

voltage relation is not influenced by altered activation of inward Na⁺ or Ca²⁺ currents under depolarization.

The neurones we used possess long neurites whose membrane potential will not be fully controlled by the voltage clamp. Because this problem will affect the amount of membrane current we record, we minimized excitation of neurites by substance P by applying the peptide close to the cell body. We also attempted to test whether lack of complete control of neurites influences the shape of the relation between substance P-sensitive membrane current and membrane potential by applying tetraethylammonium ions (TEA⁺, 50 mM), which reduce membrane potassium conductance²⁴ and may therefore be expected to alter voltage decrements along neurites. TEA⁺ reduced the size of the response to substance P, but had no effect on the shape of the current-voltage relation (Fig. 3b).

The second criterion for the dependence of the conductance on both external K⁺ and voltage was demonstrated by measuring the response in the same cell in 2.5, 12.5 and 2.5 mM K⁺ (Fig. 2). The dependence on external K⁺ concentration ([K⁺]_o) is seen more clearly if the K-chord conductance is computed (by dividing the substance P-sensitive membrane current by the driving force on K⁺) and plotted against membrane potential (Fig. 2c). The relationship between conductance and potential was shifted

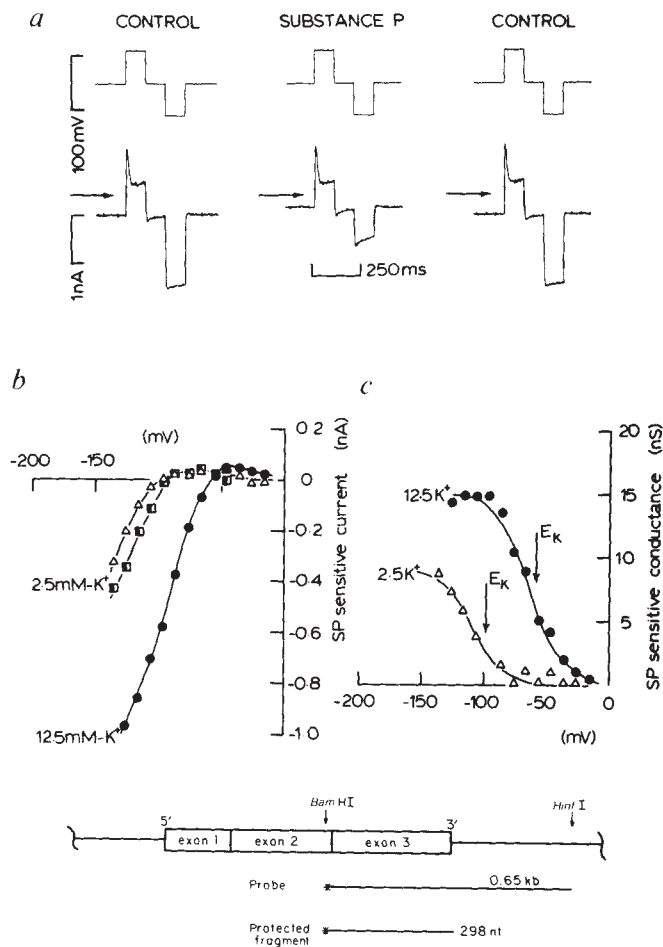


Fig. 2 a, Records of membrane potential (mp) (above) and membrane current (below) in 12.5 mM K⁺ obtained before, during and after recovery from application of 3 μ M substance P. Holding potential, -76 mV; voltage steps depolarize and hyperpolarize by 60 mV for 100 ms. The pulses were applied 100 ms apart. Horizontal arrows indicate the zero current level. Note that the A-current is activated by depolarization, but is unaffected by substance P. Inward currents are substantially reduced, and substance P seems to increase the rate of inactivation of these inward currents under hyperpolarization. Three puffs (each 1.63 s long) were applied in rapid succession to produce a brief plateau in the response to the peptide; during this time a membrane current-voltage relation was measured with pulse pairs whose amplitude increased in 10-mV steps from ± 10 to ± 60 mV. This procedure took 6 s to complete. b, Current-voltage relations from the same cell as in a, measured first in 2.5 mM K⁺ (Δ), then 12.5 mM K⁺ ($\bullet$), then in 2.5 mM K⁺ ($\square$). The current plotted is the substance P-sensitive current, measured by subtracting the current at the end of the 100-ms voltage steps applied during the action of substance P from the average of the control currents measured before and after recovery from substance P application. During the final run in 2.5 mM K⁺, no control was measured after substance P application, owing to breakdown of the seal between the recording pipette and the cell. The initial resting potential of the cell was -77 mV; diameter of cell soma, 26 μ m; temperature, 29.2 °C. In 2.5 mM K⁺ solution, NaCl concentration was 148 mM, and in 12.5 mM K⁺ it was 138 mM, otherwise the external solution had the composition given in Fig. 1 legend. Ordinate, substance P-sensitive membrane current (nA); abscissa, membrane potential (mV). c, Chord conductance for the substance P-sensitive current. The currents in 2.5 mM K⁺ (Δ) and 12.5 mM K⁺ ($\bullet$) are divided by the driving force on K⁺ ($V - E_K$, where E_K is measured from the reversal potential of the substance P-sensitive current-voltage relations). The relations are fitted by Boltzmann expressions, according to:

$$g_K = \bar{g}_K \{1 + \exp[(V - \bar{V})/k]\}^{-1}$$

where $\bar{g}_K$ is potassium conductance, $\bar{g}_K$ is limiting potassium conductance, $\bar{V}$ is the voltage at which $g_K = \bar{g}_K/2$, and k determines the steepness of the curve. For 2.5 mM K⁺ $\bar{g}_K = 9.0$ nS; $\bar{V} = -109$ mV and $k = 11.5$ mV. For 12.5 mM K⁺, $\bar{g}_K = 15.0$ nS; $\bar{V} = -62$ mV and $k = 13.8$ mV. Inward rectification of egg cells¹⁰ and skeletal muscle¹⁸ has previously been fitted to an expression similar to the above equation, with a similar steepness for voltage dependence, and an approximately similar increase in $\bar{g}_K$ with [K⁺]_o. The measured E_K (indicated by vertical arrows) shifts by 40 mV when [K⁺]_o is increased from 2.5 to 12.5 mM (theoretical, 42 mV); $\bar{V}$ shifts by 47 mV, close to the shift in E_K . Ordinate, substance P-sensitive chord conductance (nS); abscissa, membrane potential (mV).

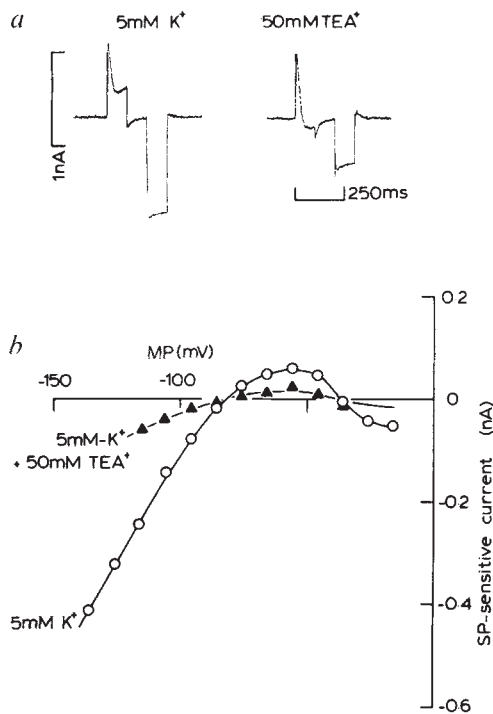
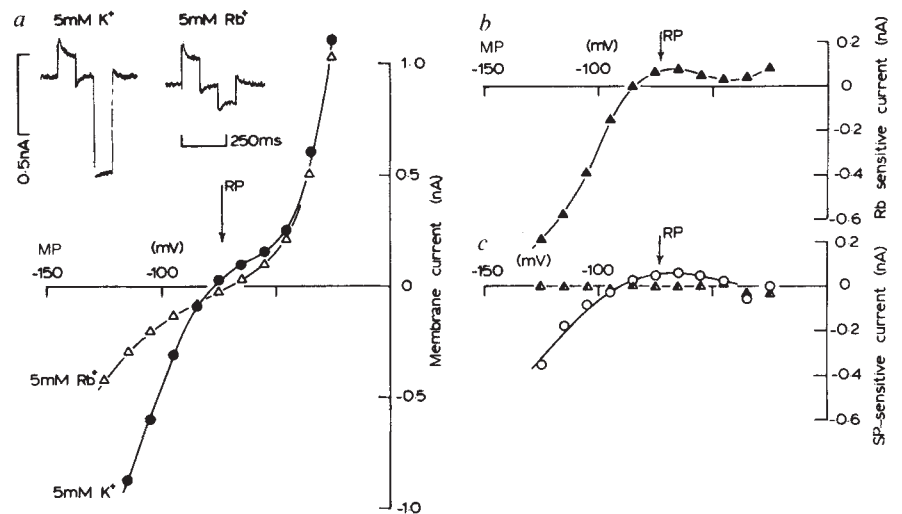


Fig. 3 *a*, Records of membrane current made under 50 mV depolarizations and hyperpolarizations from the holding potential of -76 mV in a cell immersed successively in 5 mM K^+ solution and 5 mM K^+ + 50 mM TEA^+ solution. This concentration of TEA^+ leaves the A-current virtually unaffected, but abolishes delayed K^+ currents recorded under depolarization. Inward rectification is reduced, as indicated by the reduction in inward current under hyperpolarization and by the reduced response to substance P (*b*). *b*, Substance P-sensitive current (measured as described in Fig. 2 legend, but at the start of the voltage step) plotted against membrane potential in 5 mM K^+ (○) and 5 mM K^+ + 50 mM TEA^+ solution (▲). The TEA^+ solution contained 96 mM NaCl. The line to the TEA^+ results is that to the open circles, multiplied by 0.26, indicating that the form of the current-voltage relation is not altered by TEA^+ , even though the current amplitude is reduced. Cell resting potential, -70 mV; diameter of cell soma, $24 \mu\text{m}$; temperature, 29.9°C ; diameter of substance P pipettes, $5 \mu\text{m}$; distance from cell, $12.8 \mu\text{m}$; resistance of recording pipette, $3.0 \text{ M}\Omega$.

Fig. 4 Rb^+ -induced block of inward rectification and of the effect of substance P. *a*, Membrane current-voltage relation obtained by plotting current at the beginning of the 100-ms potential step or at the peak of the A-current, which becomes activated at potentials positive to -50 mV. Currents were recorded in 5 mM K^+ (●) solution and then in 5 mM Rb^+ (zero K^+ , △) solutions. These solutions also contained NaCl at 145.5 mM, but were otherwise identical with that listed in Fig. 1 legend. The vertical arrow denotes the resting potential (RP) of the cell (-74 mV) recorded immediately after sealing the recording pipette onto the cell and breaking the membrane patch for whole cell recording. The inset shows membrane currents obtained in response to 30 mV depolarizations and hyperpolarizations from the holding potential (-76 mV) in 5 mM K^+ and 5 mM Rb^+ solutions, $3.3 \text{ M}\Omega$; diameter of cell soma, $24 \mu\text{m}$; temperature, 30.3°C . Ordinate, membrane current (nA); abscissa, membrane potential (mV). *b*, Membrane current removed by Rb^+ substitution. The current-voltage relation was obtained by subtracting the current in 5 mM Rb^+ from that in 5 mM K^+ . Arrow denotes initial resting potential of the cell. *c*, Membrane current-voltage relation for the substance P-sensitive current in 5 mM K^+ (○) and in 5 mM Rb^+ (▲) solution. These relations were obtained as in Fig. 2 by subtracting the current during substance P action from the average of currents measured in controls before and after recovery from substance P application. Vertical arrow, initial resting potential of the cell. In this cell, whole cell recording failed during the return to 5 mM K^+ solution. However, we recovered the substance P effect after Rb^+ application in one other cell.



to more positive levels at higher $[K^+]_0$, a property characteristic of inward rectification^{11,13}.

The third criterion that distinguishes inward rectification from certain other classes of K^+ conductance is its low permeability to Rb^+ (refs 10, 12, 19). Figure 4 shows that replacement of external K^+ by Rb^+ blocks inward rectification, reducing the membrane current that flows at potentials negative and slightly positive to the resting potential. Evidently, Rb^+ is itself impermeant and acts to block efflux of K^+ through the inward rectifier. The substance P-sensitive conductance of this cell had a current-voltage relation similar to that blocked by Rb^+ , and Rb^+ abolished the response to substance P (Fig. 4*b,c*).

We have investigated certain other properties of this substance P-sensitive permeability. It is blocked by the addition of external Cs^+ (2 mM), as are other inward rectifiers^{8,25-27}; it persists in the absence of external Na^+ (tetramethylammonium replacement), unlike that reported in hippocampal neurones⁹; it also persists in the absence of internal Na^+ , whose replacement with K^+ does not abolish the response to substance P. This property (and its sensitivity to TEA^+ reported above) is shared with inward rectification of skeletal muscle^{18,28}, but not with that of egg cells, which is abolished in the absence of internal Na^+ (ref. 11).

Our results with substance P on magnocellular basal forebrain neurones of the rat are reminiscent of those obtained with serotonin on *Aplysia* (molluscan) neurones by Benson and Levitan²⁵. In *Aplysia*, serotonin enhances inward rectification by phosphorylating the K^+ channels through cyclic AMP-dependent protein kinase. Here substance P reduces inward rectification. Although we have not investigated the steps between the binding of substance P to its receptor and the closure of K^+ channels, it seems likely that an intracellular second messenger system is involved²⁹, and that this is not washed out during whole cell recording for periods up to 1 h.

Substance P may have several effects on nerve cells. For example, in chromaffin cells, which secrete catecholamines, it acts indirectly, reducing the excitatory response to acetylcholine by increasing the rate of desensitization of nicotinic receptors for that transmitter³⁰. In the central cholinergic neurones used in this study, excitability was increased, the more common physiological effect of the peptide^{1,2}. As we have shown, this effect is achieved by reducing inward rectification, which contributes substantially to the resting membrane conductance (Fig. 4). Because this K^+ conductance decreases on depolariz-

ation and increases as the membrane potential becomes more negative (Fig. 2c), substance P, by reducing K^+ conductance, will induce depolarization and this depolarization will further reduce K^+ conductance. Thus, by modulating inward rectification, substance P (and serotonin in neurones of *Aplysia*) will have a self-reinforcing element to its effect on excitability.

We thank Dr Shigehiro Nakajima for help and discussion. The work was supported by NIH grant NS 10457 (Y.N.) and by a travel grant from the Wellcome Trust (P.R.S.).

Received 14 January; accepted 19 March 1985.

- Nicoll, R. A., Schenker, C. & Leeman, S. E. *Rev. Neurosci.* **3**, 227-268 (1980).
- Otsuka, M., Konishi, S., Yanagisawa, M., Tsunoo, A. & Akagi, H. *Ciba Fdn Symp.* **91**, 13-34 (1982).
- Nakajima, Y., Nakajima, S., Obata, K. & Carlson, C. G. *Soc. Neurosci. Abstr.* **10**, 659 (1984).
- Parent, A., Graves, S. & Oliver, A. *Adv. Neurol.* **24**, 1-12 (1979).
- Coyle, J. T., Price, D. L. & DeLong, M. R. *Science* **219**, 1184-1190 (1983).
- Kandel, E. R. & Tauc, L. *J. Physiol., Lond.* **183**, 287-304 (1966).
- Nelson, P. G. & Frank, K. J. *Neurophysiol.* **30**, 1097-1113 (1967).
- Constanti, A. & Galvan, M. J. *J. Physiol., Lond.* **335**, 153-178 (1983).
- Segal, M. & Barker, J. L. J. *J. Neurophysiol.* **51**, 1409-1433 (1984).
- Hagiwara, S. & Takahashi, K. *J. Membrane Biol.* **18**, 61-80 (1974).
- Hagiwara, S. & Yoshii, M. *J. Physiol., Lond.* **292**, 251-265 (1979).
- Standen, N. B. & Stanfield, P. R. *J. Physiol., Lond.* **304**, 415-435 (1980).
- Leech, C. A. & Stanfield, P. R. *J. Physiol., Lond.* **319**, 295-309 (1981).
- Marty, A. & Neher, E. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 107-122 (Plenum, New York, 1983).
- Hagiwara, S. & Ohmori, H. *J. Physiol., Lond.* **331**, 231-252 (1982).
- Katayama, Y., North, R. A. & Williams, J. T. *Proc. R. Soc. B206*, 191-208 (1979).
- Nowak, L. M. & MacDonald, R. L. *Brain Res.* **214**, 416-423 (1981).
- Hestrin, S. J. *J. Physiol., Lond.* **317**, 497-508 (1981).
- Adrian, R. H. *J. Physiol., Lond.* **175**, 134-159 (1964).
- Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 21-30 (1971).
- Gorman, A. L. F. & Thomas, M. V. *J. Physiol., Lond.* **308**, 287-313 (1980).
- Barrett, J. N., Magleby, K. L. & Pallotta, B. S. *J. Physiol., Lond.* **331**, 211-230 (1982).
- Adams, P. R., Brown, D. A. & Jones, S. W. *Br. J. Pharmacol.* **79**, 330-333 (1983).
- Stanfield, P. R. *Rev. Physiol. Biochem. Pharmacol.* **97**, 1-67 (1983).
- Benson, J. A. & Levitan, I. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3522-3525 (1983).
- Hagiwara, S., Miyazaki, S. & Rosenthal, N. P. *J. gen. Physiol.* **67**, 621-658 (1976).
- Gay, L. A. & Stanfield, P. R. *Nature* **267**, 169-170 (1977).
- Stanfield, P. R. *J. Physiol., Lond.* **209**, 231-256 (1970).
- Nishizuka, Y. *Science* **225**, 1365-1370 (1984).
- Clapham, D. E. & Neher, E. *J. Physiol., Lond.* **347**, 255-277 (1984).

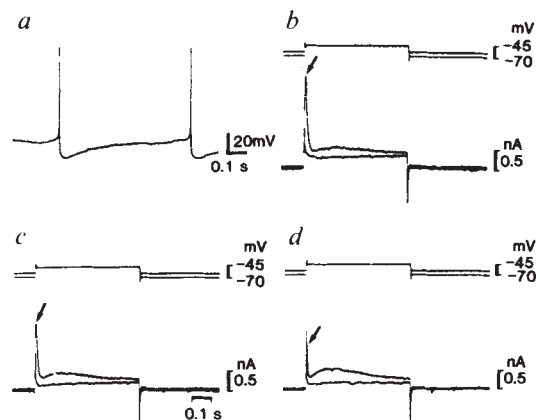


Fig. 1 A transient outward current in a serotonergic neurone and its suppression by 4-aminopyridine (4-AP). *a*, Spontaneous spikes of a typical serotonergic neurone in the dorsal raphe slice preparation; note the large AHPs, followed by a rapid phase of depolarization and then a plateau period before the subsequent spike. In *b*, the slice has been perfused with TTX (1 μ M) to eliminate sodium spikes and the cell has been voltage clamped at -70 mV (holding current, -0.01 nA); a step depolarization to -45 mV elicits a transient outward current (arrow) which peaks approximately 8 ms after the depolarizing step; a superimposed trace shows the virtual absence of a transient outward current when the holding potential is -60 mV rather than -70 mV. In *c*, 10 min after the addition of 1 mM 4-AP to the perfusate, the transient outward current (arrow) is reduced by ~30% (leakage current was estimated by subtraction of instantaneous current after the -60 to -45 mV step). In *d*, 12 min after the 4-AP has been increased to 2.5 nM, the transient outward current (arrow) is reduced by more than 60%. In *b* and *c* the current traces are beneath the voltage traces; outward currents are upward going. Photographs were taken on line from a storage oscilloscope.

Methods. Coronal midbrain slices (300-350 μ m) through the dorsal raphe nucleus were obtained from 53 male albino rats (125-200 g; Charles River). As described previously¹⁸, slices were placed on a lens paper strip at the fluid-gas interface of a slice chamber; the temperature was kept at $33 \pm 0.5^\circ\text{C}$. Artificial CSF, gassed with 95% $\text{O}_2/5\%$ CO_2 to give a final pH of 7.37-7.4, was composed of (in mM): NaCl, 130; KCl, 5.0; CaCl_2 , 2.0; MgSO_4 , 2.0; NaHCO_3 , 25.7; NaH_2PO_4 , 1.25; (+)glucose, 15. Drugs were administered in known concentrations in the perfusate; ascorbic acid (0.1 mM) was added to solutions of noradrenaline to retard oxidation. There was a continuous flow of humidified 95% $\text{O}_2/5\%$ CO_2 through the chamber. To facilitate the flow of fluid over the tissue, slices were covered with a monolayer of cotton gauze. To minimize capacitance and resistance, short stubby electrodes (7-8 mm, shank to tip; 30-50 M Ω) were pulled with a Brown-Flaming puller (Sutter Institute) using Microstar tubing (1.4 mm; Radnoti Glass Inc.); filling was with 2 M KCl. Intracellular recording and single-electrode voltage clamping were conducted using an Axoclamp-2 voltage clamp (Axon Instruments); electrode settling time was within 50-75 μ s, allowing switching frequencies of 4-6 kHz and a loop gain of at least 5 nA mV^{-1} (30% duty cycle). Although phase lag was used (to prevent oscillations), false clamping was avoided by using optimal capacitance neutralization and by selecting a switching frequency which allowed full settling to a horizontal baseline as determined by continuously monitoring input voltage¹⁹. As an operational criterion of the adequacy of the space clamp, loop gain was increased until high-threshold calcium spikes, induced by step depolarizations, were completely clamped; this was carried out in the presence of TTX (1 μ M; Sigma) to block sodium spikes. The dorsal raphe nucleus was visualized by its surface markings in incident light; serotonergic neurones within the dorsal raphe nucleus were identified by their characteristic long-duration action potentials (2 ms), high input resistance (200-400 M Ω), large AHPs, and pacemaker potentials^{11,18}.

Modulation of a transient outward current in serotonergic neurones by α_1 -adrenoceptors

G. K. Aghajanian

Departments of Psychiatry and Pharmacology, Yale University School of Medicine and the Abraham Ribicoff Research Facilities of the Connecticut Mental Health Center 34 Park St, New Haven, Connecticut 06508, USA

The excitability of various neurones in the mammalian central nervous system (CNS), ranging from motoneurones to serotonergic neurones, is enhanced by α_1 -adrenoceptor agonists¹. Excitations mediated via α_1 -adrenoceptors are associated with a slow depolarization and an increase in input resistance, probably resulting from a decrease in resting potassium conductance^{1,2}. However, the involvement of voltage-dependent transient currents in mediating α_1 excitatory effects has not been evaluated. An early transient outward current has been described which is important in regulating the frequency of repetitive firing; it is activated by depolarizing voltage steps from potentials more negative than rest and blocked by 4-aminopyridine^{3,4}. This current, which has been termed ' I_A ', was found originally in invertebrates^{3,4} and subsequently in various vertebrate neurones⁵⁻⁸. The present single-electrode voltage-clamp study demonstrates an early transient outward current (I_A) in serotonergic neurones which is suppressed by noradrenaline and the α_1 -agonist phenylephrine; a suppression of I_A may account in part for the acceleration of pacemaker activity induced by α_1 -agonists in serotonergic neurones.

Rat serotonergic dorsal raphe neurones, investigated in voltage-clamp conditions (see Fig. 1 legend), exhibited an early

transient outward current with the characteristics of I_A as originally described in molluscan neurones^{3,4}. This current could be evoked by depolarizations from holding potentials below resting potential and suppressed by 4-aminopyridine (4-AP; $n = 5$) (Fig. 1); in unclamped cells in the absence of tetrodotoxin (TTX), 4-AP (1 mM) produced a marked activation of firing

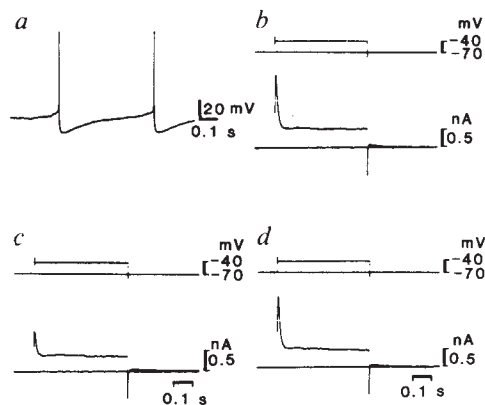


Fig. 2 Suppression by phenylephrine of the transient outward current (I_A) in a serotonergic neurone and the reversal of this effect by prazosin. *a* Shows the spikes of a neurone before clamping and treatment with TTX; this cell was not spontaneously active and spikes were evoked by 0.02 nA of depolarizing current. In *b*, TTX (1 μ M) has been added to the perfusate and the cell has been clamped at -70 mV (holding current: -0.01 nA); a step depolarization to -40 mV elicits a typical I_A . As in Fig. 1, leakage current was estimated by holding at -60 mV (where I_A is inactivated; see Fig. 3e) and measuring instantaneous current after stepping to -40 mV (trace not shown). In *c*, the amplitude of I_A is markedly reduced after perfusion (15 min) with artificial CSF containing phenylephrine (25 μ M); note that the holding current has shifted to -0.07 nA, indicating an increase in resting inward current. Although not readily discernible at this gain and holding potential, there is also a prolongation of the outward tail current (thicker trace) with little or no change in maximal amplitude during phenylephrine administration; this effect was seen consistently for both phenylephrine and noradrenaline. In *d*, 12 min after the addition of prazosin (5 μ M) in the continued presence of phenylephrine, there is a total reversal of the suppression in I_A .

rate (not shown). A further similarity to I_A is the fact that in artificial cerebrospinal fluid (CSF) containing 5 mM KCl, tail currents of this early transient outward current had a reversal potential of approximately -82 mV (± 3 s.d.; $n = 4$ cells); this shifted to -68 mV (± 3 s.d.) in 10 mM KCl, as would be predicted by the Nernst equation for a current carried by potassium ions. Based on these criteria, the early outward current of serotonergic neurones is hereafter designated I_A . To achieve rapid responses, the effects of noradrenaline or phenylephrine on the amplitude of I_A were tested at a concentration of 25 μ M, which is approximately 10 times their ED_{50} for activating the firing of serotonergic neurones¹¹. Phenylephrine (25 μ M) produced a marked reduction in the amplitude of I_A (Fig. 2). For the total group of serotonergic cells tested in this manner, phenylephrine and noradrenaline produced a 38% (± 10 s.d.; $n = 8$) and 34% (± 13 s.d., $n = 8$) reduction in I_A , respectively (leakage current was subtracted as described in Fig. 1 legend); increasing the concentration of phenylephrine or noradrenaline produced little or no further change. Activation/inactivation curves showed that suppressions of I_A by phenylephrine (Fig. 3; $n = 5$) and noradrenaline ($n = 5$) were present throughout a wide range of voltages, with initial activation occurring between -60 and -50 mV; in addition, phenylephrine reduced the leakage current (Fig. 3d), an effect which corresponds to the increase in input resistance detected by current-clamp methods¹.

The complex time course of current offset (see, for example, Fig. 3a, b) suggests the presence of transient currents in addition to I_A ; these may include rapidly activating Ca^{2+} and Ca^{2+} -dependent potassium currents (I_{Ca} and $I_{K(Ca)}$)^{9,10}. To evaluate the possible involvement of Ca^{2+} -dependent currents, the Ca^{2+} -channel blocker cadmium (1 nM) was perfused until an early inward current (arrow, Fig. 4a) and a delayed outward current (including an outward tail current; double arrows, Fig. 4a) were blocked (Fig. 4b); these currents, which were elicited from a holding potential at which I_A is inactivated, presumably rep-

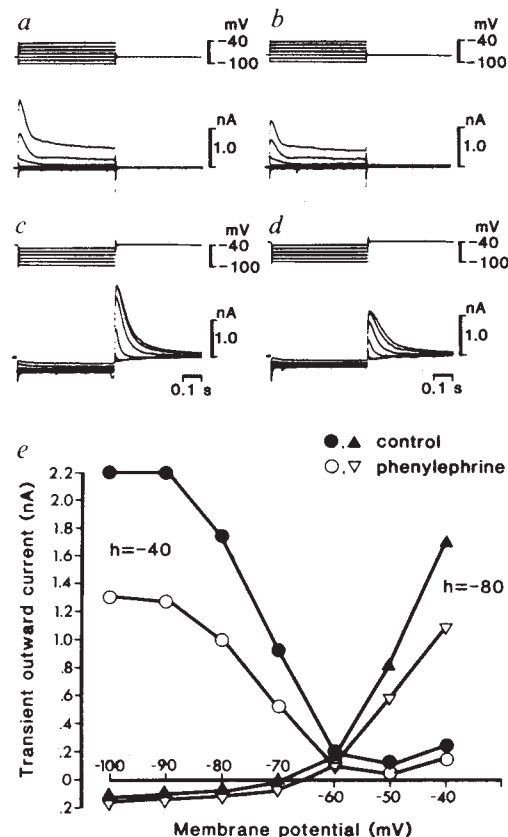


Fig. 3 The effect of phenylephrine on I_A activation/inactivation curves. In *a* and *b*, holding potential (h) is at -80 mV (holding current in *a*, -0.07 nA); current responses (lower traces) are to voltage steps (upper traces) ranging from -40 to -100 mV. Note that I_A first becomes activated at -60 mV and increases progressively up to the -40 mV step. In *b*, the administration of phenylephrine (25 μ M; 18 min) results in a decrease in amplitude at each step. In *c* and *d*, the holding potential is at -40 mV (holding current in *c*; +0.25 nA) with hyperpolarizing prepulse steps ranging from -50 to -100 mV; note that I_A is de-inactivated only by hyperpolarizing steps below -60 mV. In *d*, following the administration of phenylephrine (15 min), the holding current shifts to +0.16 nA and there is a reduction in the amplitude of I_A elicited by a return to the holding potential (-40 mV). Note the decrease in leakage current in the prepulse series (-50 to -100 mV) during phenylephrine perfusion as compared with the control series in *c*. In *e*, control and phenylephrine traces for holding potentials of $h = -40$ mV and $h = 80$ mV are plotted as direct current-voltage (I - V) curves (no leakage current subtraction); as can be seen, at all voltage steps where I_A is activated, phenylephrine reduces the maximum amplitude. Also note that phenylephrine induces a small increase in 'resting' inward current. The slice was perfused with artificial CSF containing TTX (1 μ M) throughout the experiment to block sodium spikes.

resent components of I_{Ca} and $I_{K(Ca)}$, respectively. When I_A was once again evoked, it was found to be undiminished by Cd^{2+} ; in these conditions, phenylephrine had its usual suppressant effect (Fig. 4c, d; $n = 4$).

Pharmacological studies have indicated that the noradrenergic activation of serotonergic neurones is mediated specifically by an α_1 -adrenoceptor^{11,12}. Accordingly, prazosin, a highly specific α_1 -antagonist in electrophysiological studies¹³, completely reversed the noradrenaline- or phenylephrine-induced depression of I_A (Fig. 2d; $n = 4$ and 5, respectively).

The results of the present study show that serotonergic dorsal raphe neurones possess an early transient outward current (I_A) which can be elicited by depolarizations from potentials below resting levels and is suppressed by noradrenaline or phenylephrine. As serotonergic neurones in the dorsal raphe nucleus receive a direct noradrenergic synaptic input¹⁴, the present

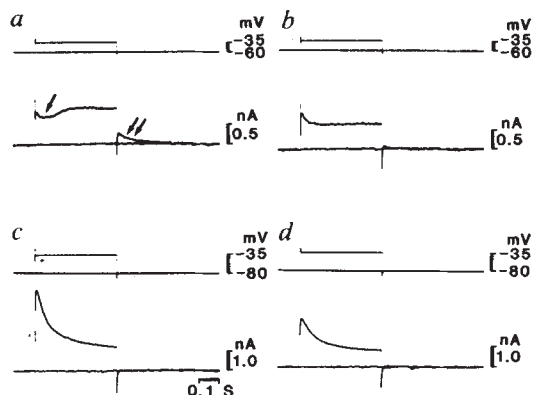


Fig. 4 Effect of the Ca^{2+} -channel blocker Cd^{2+} on the phenylephrine induced reduction of I_A . In *a*, a depolarizing step ($h = -60$ mV; holding current, $+0.02$ nA) to the command potential (-35 mV) induced a slow inward current (arrow) followed by a delayed outward current (note the outward tail current, double arrows); also note that at this holding potential I_A is largely inactivated. In *b*, after perfusion with Cd^{2+} (1 mM) for 20 min, the slow inward current and the subsequent outward currents are virtually abolished. In *c*, Cd^{2+} perfusion was continued but here the holding potential was changed to -80 mV to elicit I_A (note the change in current scale). I_A during Cd^{2+} was not diminished relative to its amplitude before Cd^{2+} (elicited from -80 mV; not shown). In *d*, phenylephrine (applied for 9 min) produces its usual reduction in I_A ; full recovery from the effect of phenylephrine occurred within a 20-min washout period. After prolonged exposures to Cd^{2+} (for example, >1 h; not shown), it was no longer feasible to test for the response to phenylephrine because of increased membrane instability. In these experiments the artificial CSF was modified by substituting equimolar HEPES for NaH_2PO_4 to prevent the precipitation of Cd^{2+} .

results suggest that the suppression of I_A by noradrenaline represents a true neurotransmitter-induced modulation of this voltage-dependent current.

Intracellular recordings both *in vivo*¹⁵ and *in vitro*¹¹ show that serotonergic neurones are tonic pacemaker cells whose spikes arise from gradual interspike depolarizations rather than discrete synaptic potentials. The interspike behaviour of unclamped serotonergic neurones is characterized by a large afterhyperpolarization (AHP), then a short period of relatively rapid depolarization, and finally a plateau region lasting several hundred milliseconds which precedes spike initiation¹¹. α_1 Stimulation markedly shortens the plateau period but does not reduce the initial amplitude of the AHP¹¹. Similarly, in preliminary voltage-clamp experiments, the amplitude of $I_{K(\text{Ca})}$ was not reduced and, in fact, its duration was increased by phenylephrine and noradrenaline (Fig. 2c legend). This contrasts with β -adrenoceptor effects on hippocampal pyramidal neurones which are characterized by an inhibition of the Ca^{2+} -dependent AHP, resulting in increased excitability due to reduced spike frequency accommodation^{16,17}. The plateau period of the pacemaker potential in serotonergic neurones occurs between -60 to -50 mV; this is precisely the range in which I_A becomes activated (Fig. 3e). Thus, by suppressing I_A , α_1 -adrenoceptors would abbreviate this plateau period and thereby, together with a reduction in resting potassium conductance¹, accelerate pacemaker activity. It remains to be determined whether increases in neuronal excitability induced by α_1 -receptors elsewhere in the CNS also involve a suppression of I_A .

I thank Dr C. F. Stevens for helpful discussion and N. Margiotta for technical assistance. This work was supported by PHS grant MH 17871 and the State of Connecticut.

- VanderMaelen, C. P. & Aghajanian, G. K. *Nature* **287**, 346–347 (1980).
- Connor, J. A. & Stevens, C. F. *J. Physiol., Lond.* **213**, 31–55 (1971).
- Adams, D. J., Smith, S. J. & Thompson, S. H. *A. Rev. Neurosci.* **3**, 141–167 (1980).
- Adams, P. R., Constanti, A. & Brown, D. A. *J. Physiol., Lond.* **330**, 537–572 (1982).
- Galvan, M. *Neurosci. Lett.* **31**, 295–300 (1982).
- Gustafsson, B., Galvan, M., Grafe, P. & Wigstrom, H. *Nature* **299**, 252–254 (1982).
- Segal, M., Rogawski, M. A. & Barker, J. L. *J. Neurosci.* **4**, 604–609 (1984).
- Adams, P. R., Constanti, A., Brown, D. A. & Clark, R. B. *Nature* **296**, 746–749 (1982).
- MacDermott, A. B. & Weight, F. F. *Nature* **300**, 185–188 (1982).
- VanderMaelen, C. P. & Aghajanian, G. K. *Brain Res.* **289**, 109–119 (1983).
- Baraban, J. M. & Aghajanian, G. K. *Neuropharmacology* **19**, 355–363 (1980).
- Menkes, D. M., Baraban, J. M. & Aghajanian, G. K. *Naunyn-Schmiedeberg Archs Pharmacol.* **317**, 273–275 (1981).
- Baraban, J. M. & Aghajanian, G. K. *Brain Res.* **204**, 1–11 (1981).
- Aghajanian, G. K. & VanderMaelen, C. P. *Brain Res.* **238**, 463–469 (1982).
- Madison, D. V. & Nicoll, R. A. *Nature* **299**, 636–638 (1982).
- Haas, H. L. & Konnerth, A. *Nature* **302**, 432–434 (1983).
- Aghajanian, G. K. & Lakoski, J. M. *Brain Res.* **305**, 181–185 (1984).
- Finkel, A. S. & Redman, S. J. *Neurosci. Meth.* **11**, 101–127 (1984).

Modulation of single Ca^{2+} -dependent K^+ -channel activity by protein phosphorylation

Douglas A. Ewald, Alan Williams* & Irwin B. Levitan†

Graduate Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254, USA

There is considerable evidence that cyclic AMP can modulate the electrical activity of excitable cells^{1–10} and that protein phosphorylation by the catalytic subunit (CS) of cAMP-dependent protein kinase is a necessary step in these modulatory effects^{11–21}. In analogy to alterations in enzyme activities following phosphorylation¹², it seems possible that direct phosphorylation of ion-channel proteins may alter their gating properties, giving rise to the observed changes in electrical activity. However, the results obtained so far do not indicate whether it is ion channels themselves that are phosphorylated, or whether phosphorylation is simply an early step in some cascade of events which leads ultimately to modulation of channel activity. The development of single-channel recording techniques²² has provided a way to investigate this question. Here we describe effects of CS on the activity of individual Ca^{2+} -dependent K^+ channels from the nervous system of the land snail *Helix* measured in isolated membrane patches and in artificial phospholipid bilayers. The results demonstrate that cAMP-dependent protein phosphorylation produces long-lasting changes in the activity of individual channels, and indicate that the relevant phosphorylation site is closely associated with the channel.

Helix ganglia contain an identified group of neurones (the F cluster) in which the outward current is predominantly Ca^{2+} -dependent²³, and DePeyer *et al.*¹⁹ found that intracellular perfusion of CS could selectively enhance this Ca^{2+} -dependent K^+ current in such neurones. Using detached patches from these neurones in the 'inside-out' configuration²², we have observed Ca^{2+} -dependent K^+ channels with Ca^{2+} and voltage dependences as shown in Fig. 1. The maximal single channel conductance, measured using symmetrical K^+ solutions in the detached patch configuration, is 40–60 pS. Thus, this channel is distinct from the large conductance (200–300 pS) Ca^{2+} -dependent K^+ channels which have been described in several systems²⁴. Most of our experiments have been performed using 1 mM K^+ in the electrode (the outside of the membrane) and 120 mM K^+ in the solution bathing the inner membrane surface, to approximate the K^+ gradient seen by the channel in the normal cell membrane. In these conditions the apparent single-channel conductance measured at 0 mV is ~ 20 pS, similar to that reported by Lux *et al.*²⁵ for this channel in cell attached patches. Furthermore, the channel is K^+ selective, and thus is distinct from a Ca^{2+} -dependent non-selective cation channel which we have observed in *Helix* neurones (D.A.E. and I.B.L., unpublished) and which has been described previously in several cell types^{26–28}.

Received 19 November 1984; accepted 13 March 1985.

1. Aghajanian, G. K. in *Catecholamines: Neuropharmacology and Central Nervous System—Theoretical Aspects* (eds Usdin, E., Carlsson, A., Dahlstrom, A. & Engel, J.) 85–92 (Liss, New York, 1984).

* Permanent address: Cardiothoracic Institute, 2 Beaumont Street, London W1N 2DX, UK.
† To whom correspondence should be addressed.

Fig. 1 A Ca^{2+} -dependent K^+ channel in an isolated patch of *Helix aspersa* neurone membrane. **a**, Continuous record of channel activity at 0 mV in a patch with its inner surface exposed to 10^{-7} M free Ca^{2+} . There are several channels in this patch (the dotted lines and the numbers to the right of the traces indicate the number of open channels). The single channel current is ~ 2.2 pA at the Na^+ plus K^+ reversal potential, and approaches zero at the K^+ reversal potential, indicating that the channels are K^+ selective. **b**, The same patch exposed to 10^{-8} M free Ca^{2+} . Channel activity is decreased substantially by this 10-fold decrease in free Ca^{2+} , indicating that the channels are Ca^{2+} dependent. There is no indication of any activity of other types of channels in this patch. In the experimental conditions described below, we often obtained patches such as this where the activity over a wide potential range was entirely due to this type of channel. Channel activity in such patches was analysed by integrating the activity over baseline for a certain length of time and dividing by the integral of one channel being open for that length of time. Thus, channel activity is expressed in units of Np where N is the number of active channels in the patch and p is the probability of opening of a single channel. **c**, Analysis of channel activity in this patch as a function of membrane voltage for both 10^{-7} M ($\square$) and 10^{-8} M (Δ) free Ca^{2+} . The data were fitted by eye with the approximation to a Boltzmann distribution given by $(Np)_{\text{max}}(e^{(V-V_{\text{mid}})/\text{slope}}/1 + e^{(V-V_{\text{mid}})/\text{slope}})$ where V is membrane voltage and V_{mid} is the midpoint of the curve along the voltage axis. The fitted curves for 10^{-7} M and 10^{-8} M free Ca^{2+} had virtually the same maximum Np (1.25 and 1.30 respectively) and slope (e -fold change per 20.7 mV and 23.6 mV respectively), but the midpoints of the curves were about 30 mV apart. This shift is typical of our results in other experiments. **Methods.** The *Helix* central nervous system was desheathed and decapsulated in *Helix* saline containing 80 mM NaCl, 4 mM KCl, 10 mM CaCl_2 , 5 mM MgCl_2 , 10 mM glucose, and 5 mM HEPES (pH 7.4). The exposed neurones were then bathed for 50 min at 37°C in *Helix* saline containing 10 mg ml^{-1} trypsin. After 25 min the trypsin solution was taken up in a Pasteur pipette and gently flushed over the preparation for 2 min. Enzyme treatment was terminated with six rinses of 37°C saline. Standard patch clamp recording techniques²² were used to form 'inside-out' patches with seal resistances of 10 G Ω or more from neurones in the F cluster. The patch pipette solution contained 60 mM NaCl, 1 mM KCl, 40 mM CaCl_2 , 5 mM MgCl_2 and 5 mM HEPES (pH 7.4). The inner surface of the membrane was bathed in EGTA-buffered Ca^{2+} solutions containing 5 mM EGTA, 120 mM KCl, 1 mM MgCl_2 , 0.25 mM ATP and 5 mM HEPES (pH 7.4). The total concentration of CaCl_2 was either 1.13 mM or 3.72 mM to give free Ca^{2+} concentrations of $\sim 10^{-8}$ M and 10^{-7} M, respectively. The K^+ and Na^+ concentrations in these pipette and bath solutions correspond to a K^+ reversal potential of -120 mV and a Na^+ plus K^+ reversal potential of -20 mV.

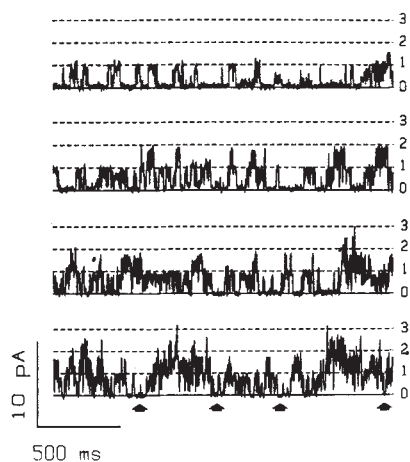
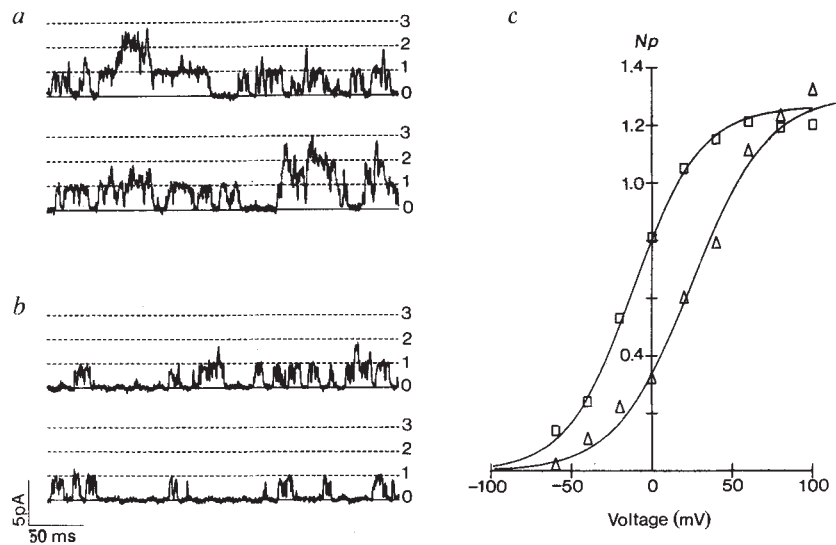


Fig. 2 Modulation of Ca^{2+} -dependent K^+ channel activity in an isolated membrane patch by catalytic subunit of cAMP-dependent protein kinase (CS). CS was purified to homogeneity by the method of Peters *et al.*³⁴ and stored at -20°C . The preparation was diluted 100-fold with 10^{-8} M free Ca^{2+} bath solution immediately before use. The patch shown here was held at 0 mV in 10^{-8} M free Ca^{2+} and contained at least five Ca^{2+} -dependent K^+ channels. This continuous record of channel activity begins 50 s after exchanging the 0.2-ml chamber with 0.4 ml of 10^{-8} M free Ca^{2+} bath solution containing 0.4 μM CS. Channel activity increased abruptly during the 8 s of this record and remained at this heightened level for the remainder of the patch lifetime (5 min). Arrows below the bottom trace indicate examples of occasional time periods when all channels are closed amidst this heightened activity.

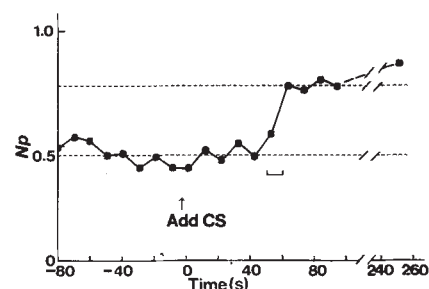


Fig. 3 Time course of the increase in Ca^{2+} -dependent K^+ channel activity produced by CS. Each point is the Np value for a 10.24-s interval during the 80 s before and 100 s after the addition of 0.4 μM CS (vertical arrow). The patch was held at 0 mV during this time. The traces in Fig. 2 were taken at the time indicated by the bracket, during the abrupt increase in channel activity resulting from CS addition. Dashed lines are drawn at the average Np levels before ($Np = 0.45$) and after ($Np = 0.76$) the abrupt increase. At 100 s, a set of voltage steps was begun which ended at 250 s, and the Np value at 0 mV between 250 and 260 s is also shown. In the six other experiments performed in our laboratory, the onset of the effect was similarly abrupt and in six out of seven cases the increase in channel activity lasted for at least 5 min. In four additional experiments performed in collaboration with J. DePeyer and H. Reuter, similar abrupt and long-lasting increases in the activity of this channel were seen. In all three experiments in which patches with this channel were exposed to heat-inactivated CS, no changes in channel activity were seen for at least 10 min. In one of these experiments the patch was subsequently exposed to active CS and there was an abrupt increase in channel activity within 4 min.

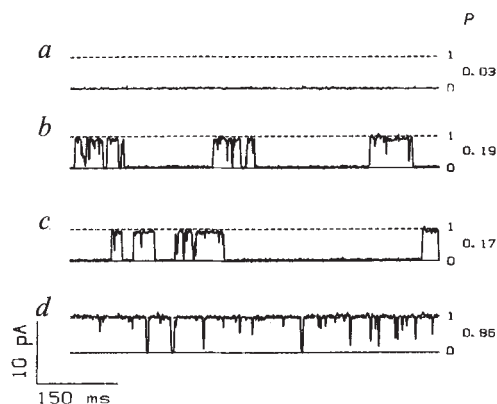


Fig. 4 Modulation of a reconstituted Ca^{2+} -dependent K^{+} channel in a phospholipid bilayer formed on the tip of a patch electrode. A single channel which had been observed immediately after formation of the bilayer remained closed virtually all the time in a well containing 1.13 mM CaCl_2 to give a free Ca^{2+} concentration of about 10^{-8} M (a). When the electrode was moved to a well containing about 10^{-6} M free Ca^{2+} (4.84 mM CaCl_2), channel openings and closings were observed (b). When the bath solution contained either more or less KCl than the electrode, the single channel currents reversed at the calculated K^{+} equilibrium potential (-17 mV in this experiment); furthermore addition of Na^{+} to either the bath or the electrode solution had no effect on the reversal potential, indicating that the channel is K^{+} selective. In some experiments the channel exhibited openings even at free Ca^{2+} concentrations as low as 1 nM, whereas in others it opened only in the presence of $\geq 10 \mu\text{M}$ free Ca^{2+} . Furthermore, this Ca^{2+} -dependent K^{+} channel was only observed in about 30% of our membrane preparations. We cannot explain this variability between experiments. However, the effects of protein phosphorylation on the activity of this channel (in this case in the presence of $\sim 10^{-6}$ M free Ca^{2+}) were reliable and consistent from experiment to experiment. Treatment with $0.7 \mu\text{M}$ DTNB-inactivated CS^{34} for 13 min had no effect on channel open probability (c). In contrast, treatment with $0.1 \mu\text{M}$ active CS led to a dramatic increase in open probability (d). The values to the right of each trace are channel open probabilities for each experimental condition, calculated from at least 30 s of record containing many opening and closing transitions. The voltage across the bilayer was $+40$ mV (corresponding to depolarization of the cell membrane) for the traces shown here.

Methods. F clusters were dissected from *Helix* central nervous system, and were rinsed sequentially with Listerine, then with 10 mM EGTA (pH 7.4), and finally with 20 mM HEPES (pH 7.4). They were then homogenized in the latter buffer (clusters from five *Helix* per ml of homogenizing buffer) using five strokes of a glass-glass homogenizer. The homogenate was centrifuged at 10,000g for 15 min and the pellet was discarded. The supernatant was centrifuged at 100,000g for 60 min; the resulting supernatant was discarded and the pellet was taken as the source of membranes. The pellet was resuspended in 200 μl of 200 mM KCl, 20 mM HEPES (pH 7.4) by homogenization. To this was added 200 μl of a mixture of brain phosphatidylethanolamine/phosphatidylserine (PE/PS, 3/1; Avanti) which had been suspended in 20 mM HEPES and sonicated until the suspension was clear. CaCl_2 was added to a final concentration of 0.1 mM, and the membrane/phospholipid mixture was frozen and thawed three times. The reconstituted preparation contained, in a total volume of 400 μl , 300–500 μg membrane protein ml^{-1} , 4 mg phospholipid ml^{-1} , 20 mM HEPES, 100 mM KCl and 0.1 mM Ca^{2+} . This preparation (100 μl) was placed in a well of a sterile microtitre plate (Falcon). A fresh plate was used for each experiment and all solutions were filtered through $0.45 \mu\text{m}$ filters to avoid spurious channels arising from bacterial contamination. The preparation was left for at least 5 min at room temperature to allow a phospholipid monolayer to form at the air/liquid interface. A fire polished patch electrode containing 50 mM KCl, 20 mM HEPES (pH 7.4) was pulled through the monolayer twice to form a bilayer on the electrode tip^{32,33}. The seal resistance for this experiment was 17 G Ω , and the bandwidth was 1 kHz. Current across the bilayer was monitored, and when single-channel currents were detected the electrode was moved to an adjacent chamber containing 100 mM KCl, 20 mM HEPES (pH 7.4), 1 mM MgCl_2 , 0.25 mM ATP, 5 mM EGTA, and CaCl_2 and enzymes as described above.

Figure 1c shows the voltage dependence of channel opening at two different Ca^{2+} concentrations. Because we never obtain patches which contain only one Ca^{2+} -dependent K^{+} channel, channel activity is expressed in units of Np , where N is the number of active channels in the patch and p is the open probability for an individual channel. Note that increasing the Ca^{2+} concentration shifts the Np -voltage curve to the left, with little change in slope or in the maximum Np value. The shift in the mid-point of the curve is ~ 30 mV for a 10-fold change in Ca^{2+} concentration.

We have investigated the possibility that the activity of this channel can be regulated by cAMP-dependent protein phosphorylation, by applying CS together with Mg^{2+} and ATP directly to the inner membrane surface of inside-out patches. The results of one such experiment are illustrated in Fig. 2. At the beginning of the record (top trace) the channel activity is that normally observed in the presence of 10^{-8} M Ca^{2+} at this voltage (0 mV, compare with Fig. 1b). During the 8-s period shown, channel activity increases dramatically, so that by the end of the record (bottom trace) several channels are open a large portion of the time.

Figure 3 shows the time course of the onset of this response. It can be seen that the Np value was stable before and for almost 1 min after addition of CS (vertical arrow), and then abruptly increased to reach a new stable level. Such abrupt increases in activity were observed in all experiments in which active CS was added ($n=7$), with a latency ranging from 12 s to 4 min. In all but one experiment the activity remained elevated for many minutes, often for as long as we could hold the patches, although on occasion there did seem to be some reversal with time. In contrast to these results with active CS, we have never seen such changes in channel activity either spontaneously, or following addition of heat-inactivated CS.

A consistent finding in all our experiments is that CS causes an increase in Np at all voltages between -60 and $+100$ mV. Although the reasons are unclear, it seems likely that CS is having more than one effect. The results are consistent with the possibility that phosphorylation leads to an increase in the affinity of Ca^{2+} -dependent K^{+} channels for Ca^{2+} , which shifts the Np -voltage curve to the left at negative voltages. The maximum Np at positive voltages is also increased, which is consistent with an increase in the number of functional channels in the patch, or with an increase in the maximum open probability for a given channel.

These and other isolated patch experiments^{29,30} have provided direct evidence that the activity of individual channels can be modulated by cAMP-dependent protein phosphorylation, and demonstrate that the phosphorylation target must be something which remains associated with the patch of membrane when it is detached from the cell. This target may be the channel itself, or it may be some other membrane-associated element (for example, a cytoskeletal component) which comes away with the patch. To study this further, we have examined channel activity in a reconstituted system using homogenized F cluster neurones dissected from *Helix* ganglia and a crude membrane vesicle fraction prepared by differential centrifugation. When these vesicles are allowed to fuse with artificial phospholipid bilayers in conditions which permit the measurement of single channel currents across the bilayer³¹, we find at least five separate K^{+} channels which can be distinguished on the basis of their kinetics and single channel conductances (I.B.L. and A.W., in preparation). At least one of these K^{+} channels is Ca^{2+} -dependent. These experiments have been performed in several different configurations, including large planar bilayers (80–200 μm diameter) formed across an opening between two chambers³¹, and in bilayers formed on the tip of a patch electrode ('tip-dipping')^{32,33}. Figure 4 illustrates an experiment in which a single Ca^{2+} -dependent K^{+} channel was present in the artificial membrane on the tip of a patch pipette. This channel has a single-channel conductance of 70–100 pS, somewhat larger than that of the Ca^{2+} -dependent K^{+} channel observed in membrane patches. It also seems to be less sensitive to Ca^{2+} , although the

Ca^{2+} sensitivity of the bilayer channel varies markedly between experiments. It remains to be determined whether the channels in the patch and the bilayer are distinct, or whether the apparent differences in some of their properties result simply from differences in their membrane environment.

The bilayer channel is closed virtually all of the time in $10^{-8} \text{ M Ca}^{2+}$ at +40 mV (Fig. 4a), but is open for a significant time when the Ca^{2+} concentration is raised to 10^{-6} M (Fig. 4b). This degree of Ca^{2+} sensitivity is in the middle of the wide range that we observed. There is no effect on channel activity after prolonged treatment (in this case 13 min) with dithionitrobenzene (DTNB)-inactivated or heat-inactivated CS (Fig. 4c). However, the opening and closing kinetics of the channel change dramatically within 20 s of the addition of active CS, with a large increase in open probability (Fig. 4d). The latency for the onset of the response ranged from too fast to measure (that is, <5 s) to as long as 7 min. The p values to the right of each trace in Fig. 4 demonstrate the magnitude of the probability change at +40 mV produced by phosphorylation with active CS; such changes were observed in all the experiments ($n=11$) carried out with either the large planar bilayers or the 'tip-dip' technique. Investigation of the change in probability elicited at various voltages by CS treatment produced results (not shown) entirely consistent with a shift in channel affinity for Ca^{2+} .

Previous studies from our own and other laboratories have provided evidence that protein phosphorylation can regulate the membrane properties of excitable cells. The present results extend these findings by demonstrating directly that the activity of individual Ca^{2+} -dependent K^{+} channels can be modulated by cAMP-dependent protein phosphorylation. In addition, the results put the phosphorylation site considerably closer to the channel; indeed, from the reconstitution experiments, in which individual channels are effectively at infinite dilution in the bilayer, we conclude that the phosphorylation target is either the K^{+} channel itself or something intimately associated with the channel. Finally, clues have been provided about those aspects of channel function which are modulated by phosphorylation, and it will be of great interest to extend these approaches to investigate the detailed molecular mechanisms of ion-channel regulation.

We thank Dr Christopher Miller for advice on bilayer strategy and tactics, and Drs Miller, Gary Yellen and John Lisman for their critical comments on the manuscript. Several early phosphorylation experiments in the isolated patch were done with Drs Harald Reuter and Jacques DePeyer at the University of Bern; we thank them for their collaboration and for teaching us how to patch clamp. This work was supported by NIH grant NS17910 to I.B.L.

Received 19 November 1984; accepted 12 March 1985.

1. Tsien, R. W. *Nature new Biol.* **245**, 120-122 (1973).
2. Reuter, H. *J. Physiol., Lond.* **242**, 429-451 (1974).
3. Treisman, S. & Levitan, I. B. *Nature* **261**, 62-64 (1976).
4. Kaczmarek, L., Jennings, K. & Strumwasser, F. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5200-5204 (1978).
5. Drummond, A., Benson, J. & Levitan, I. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5013-5017 (1980).
6. Deterre, P., Paupardin-Tritsch, D., Bockaert, J. & Gerschenfeld, H. *Nature* **290**, 783-785 (1981).
7. Klein, M., Camardo, J. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5713-5717 (1982).
8. Siegelbaum, S., Camardo, J. & Kandel, E. R. *Nature* **299**, 413-417 (1982).
9. Lemos, J. & Levitan, I. B. *J. gen. Physiol.* **83**, 269-285 (1984).
10. Siegelbaum, S. & Tsien, R. W. *Trends Neurosci.* **6**, 307-313 (1983).
11. Kuo, J. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1349-1355 (1969).
12. Glass, D. & Krebs, E. *Ann. Rev. Pharmac. Tox.* **20**, 363-388 (1980).
13. Walsh, D. & Ashby, C. *Rec. Prog. Horm. Res.* **29**, 329-359 (1973).
14. Adams, W. B. & Levitan, I. B. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3877-3880 (1982).
15. Castellucci, V., Nairn, A., Greengard, P., Schwartz, J. & Kandel, E. R. *J. Neurosci.* **2**, 1673-1681 (1982).
16. Osterrieder, W. *et al. Nature* **298**, 576-578 (1982).
17. Kaczmarek, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7487-7491 (1980).
18. Castellucci, V. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7492-7496 (1980).
19. DePeyer, J., Cachelin, A., Levitan, I. B. & Reuter, H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4207-4211 (1982).
20. Alkon, D., Acosta-Urquidí, J., Olds, J., Kuzma, G. & Neary, J. *Science* **219**, 303-306 (1983).
21. Levitan, I. B., Lemos, J. & Novak-Hofer, I. *Trends Neurosci.* **6**, 496-499 (1983).
22. Hamill, O., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
23. Heyer, C. & Lux, H. D. *J. Physiol., Lond.* **262**, 349-382 (1976).
24. Latorre, R. & Miller, C. *J. Membrane Biol.* **71**, 11-30 (1983).
25. Lux, H. D., Neher, E. & Marty, A. *Pflügers Arch. ges. Physiol.* **389**, 293-295 (1981).

26. Colquhoun, D., Neher, E., Reuter, H. & Stevens, C. *Nature* **294**, 752-754 (1981).
27. Yellen, G. *Nature* **296**, 357-359 (1982).
28. Maruyama, Y. & Petersen, O. *Nature* **300**, 61-63 (1982).
29. Shuster, M., Camardo, J., Siegelbaum, S., Eppler, C. & Kandel, E. R. *Soc. Neurosci. Abstr.* **10**, 145 (1984).
30. Ewald, D. & Levitan, I. B. *Soc. Neurosci. Abstr.* **10**, 145 (1984).
31. Miller, C. *Physiol. Rev.* **63**, 1209-1242 (1983).
32. Suarez-Isla, B., Wan, K., Lindstrom, J. & Montal, M. *Biochemistry* **22**, 2319-2323 (1983).
33. Coronado, R. & Latorre, R. *Biophys. J.* **43**, 231-236 (1983).
34. Peters, K., Demaille, J. & Fischer, E. *Biochemistry* **16**, 5691-5697 (1977).

Drastic change in idiotypic but not antigen-binding specificity of an antibody by a single amino-acid substitution

Andreas Radbruch, Siegfried Zaiss, Claudia Kappen, Marianne Brüggemann, Konrad Beyreuther & Klaus Rajewsky

Institute for Genetics, University of Cologne, Weyertal 121, D-5000 Köln 41, FRG

In proliferating B lymphocytes, somatic mutation of rearranged antibody variable (V)-region genes occurs at high frequency and may have a key role in the selection of these cells¹⁻³. It is of interest in this context to learn in which way single mutations can affect antigen binding and/or idiotypic specificity of an antibody. Previous investigations have analysed spontaneous mutants of myeloma and hybridoma cells in which the mutation affected the antigen-binding specificity of the antibody⁴⁻⁸. Here we describe an antibody mutant that has fully retained antigen-binding specificity but has lost or drastically changed all V-region antigenic determinants (idiotopes) of the wild type as defined by monoclonal anti-idiotope antibodies. The mutant phenotype is generated by a glycine to arginine exchange in the middle of the diversity (D) element, at position 103 of the heavy chain.

The mutant was derived from the monoclonal antibody B1-8.81, which binds the hapten 4-hydroxy-3-nitro-5-iodophenyl-acetyl (NIP) with high affinity^{9,10}. From a series of monoclonal anti-idiotope antibodies raised against antibody B1-8 (refs 11, 12), two were used for mutant selection. Antibody Ac38 defines an idiotope (idiotope Ac38) outside the hapten-binding cleft. In contrast, idiotope Ac146, recognized by antibody Ac146, is presumably located close to the hapten-binding site because hapten and antibody Ac146 compete for binding to the B1-8 V region. Spontaneously arising idiotope loss mutants of B1-8.81 were isolated by fluorescence-activated cell sorting. It has been shown previously that an Ac146⁻Ac38⁺ mutant has lost hapten-binding specificity together with the binding site-related idiotope Ac146⁶. The present experiments describe a mutant selected for the expression of the latter idiotope and against expression of idiotope Ac38.

Table 1 Affinity and fine specificity of hapten binding to wild-type and mutant antibody

	IC ₅₀ NIP-cap (M)	IC ₅₀ NP-cap (M)	Fine specificity
B1-8.81	1×10^{-7}	2×10^{-5}	NNP = NIP > NCLP = NP
B1-8.81V3	1.1×10^{-7}	1.8×10^{-5}	NNP = NIP > NCLP = NP

The IC₅₀ value is the molar hapten concentration resulting in 50% inhibition of binding of antibody ($60 \mu\text{g ml}^{-1}$) to N¹²⁵I-NIP-cap ($4 \times 10^{-8} \text{ M}$), determined by Farr assay as described in ref. 9. For unknown reasons, IC₅₀ values for NP-cap differed reproducibly by an order of magnitude from those published previously¹⁰. The order of haptens according to affinity was: NNP, (4-hydroxy-3,5-dinitrophenyl)acetyl; NIP, (4-hydroxy-5-iodo-3-nitrophenyl)acetyl; NCLP, (4-hydroxy-5-chloro-3-nitrophenyl)acetyl; NP, (4-hydroxy-3-nitrophenyl)acetyl.

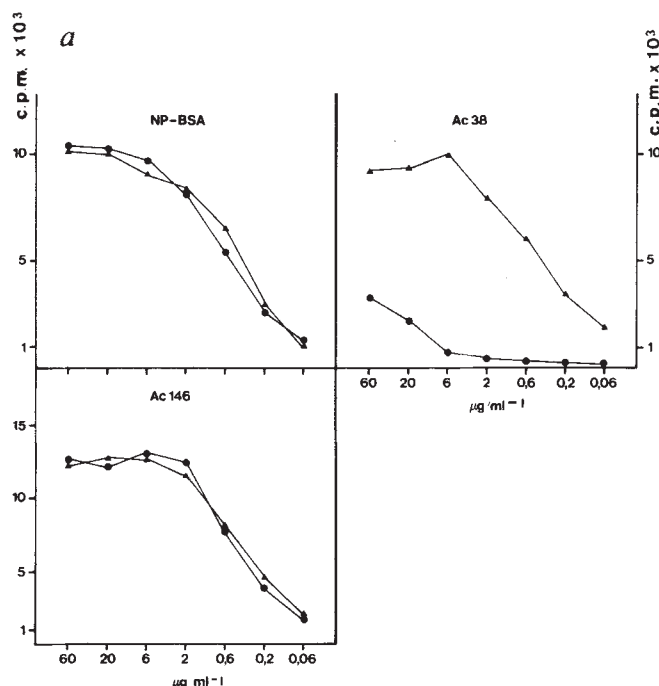
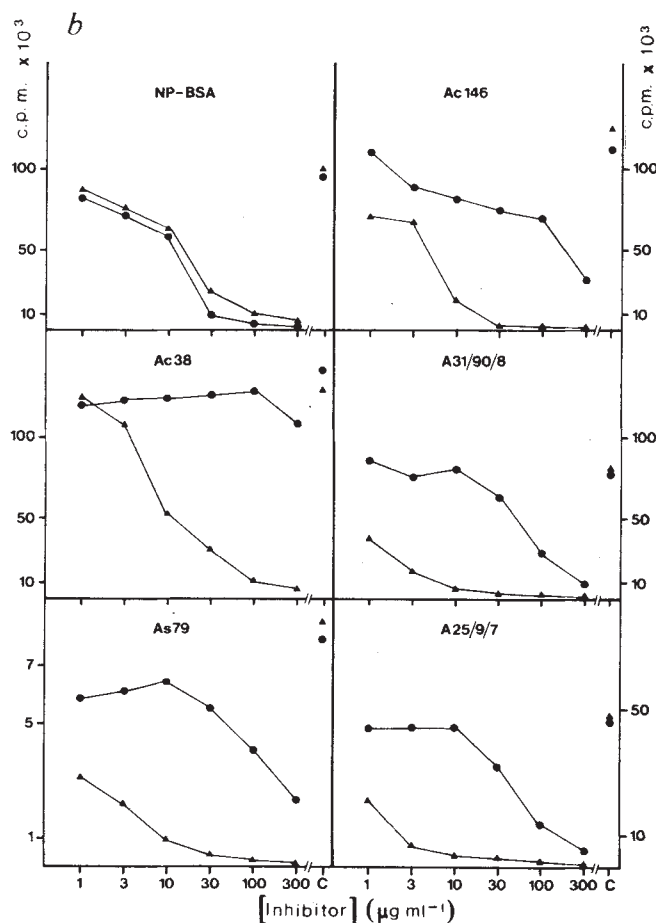


Fig. 1 Serological analysis of B1-8.δ1 (▲) and B1-8.δ1V3 (●). **a**, Direct binding in a solid-phase radioimmunoassay²⁶. Plastic plates were coated with NIP-BSA or anti-idiotope antibody at a concentration of 50 $\mu\text{g ml}^{-1}$. After saturating the plastic surface with BSA, B1-8.δ1 or B1-8.δ1V3 antibody was added to the coated wells at the concentrations indicated on the abscissa, and bound antibody detected with ^{125}I -labelled monoclonal anti-mouse $\lambda 1$ -chain antibody Ls136 (0.5 $\mu\text{g ml}^{-1}$; refs 11, 27). Results are shown for NIP-BSA and the anti-idiotopes Ac146 and Ac38. The anti-idiotope antibodies Ac106, As79, A25.9.7 and A31.90.8 behaved like antibody Ac38 in that they failed to bind the mutant B1-8.δ1V3 (data not shown). **b**, Binding inhibition. The binding of iodinated antibody B1-8.δ1 to NIP-BSA or anti-idiotope-coated plates was inhibited by graded concentrations (abscissa) of unlabelled antibody B1-8.δ1 (▲) or B1-8.δ1V3 (●). Control binding without inhibitor is indicated (C).



B1-8.δ1 cells were incubated with fluoresceinated antibody Ac146 (50 $\mu\text{g ml}^{-1}$) in the presence of a 100-fold excess of unlabelled antibody Ac38. In these conditions, only cells that have retained the expression of idiotope Ac146 but lost that of idiotope Ac38 are stained. Stained cells were then enriched in eight successive rounds of sorting, expanding the separated cells after each sort. The mutant cells were finally isolated by cloning and identified by a radioimmunoassay of culture supernatant. Fifteen clones produced antibodies of mutant phenotype, that is, antibodies binding to antibody Ac146 but not to antibody Ac38. The frequency of the variant cells in the original B1-8.δ1 population is presumably $>2 \times 10^{-8}$, as about 5×10^7 cells were used in the initial sorts. On the other hand, Ac146⁺Ac38⁻ cells were not detected in the original population by fluorescence microscopy, staining the cells simultaneously with the two anti-idiotope antibodies coupled to different fluorochromes⁴ and repeatedly screening 1×10^5 cells. These results are consistent with previous data⁴ and indicate that the frequency of a given idiotype variant in B1-8.δ1 cells is at least one order of magnitude lower than the frequency of variants of the myeloma S107 which have lost the property of antigen binding.

Because of the rare occurrence of the variants and the fact that three of the variant cell clones behaved identically when tested with a variety of other anti-idiotopes, it seemed unlikely that we had isolated a series of independent mutants. Therefore, we chose only one clone, B1-8.δ1V3, for a detailed characterization. The antibody produced by this clone (antibody B1-8.δ1V3) was assayed for binding to a NIP-conjugate of bovine serum albumin (NIP₁₄-BSA) and the anti-idiotope antibodies

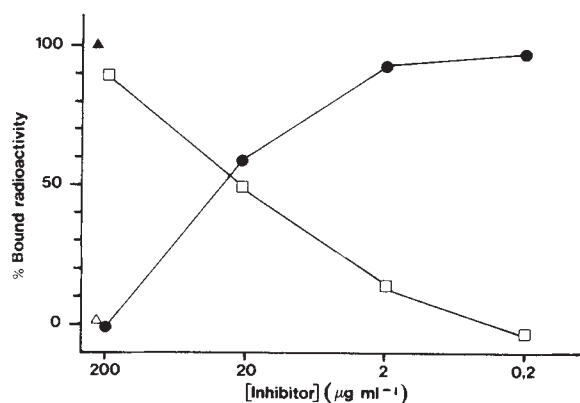
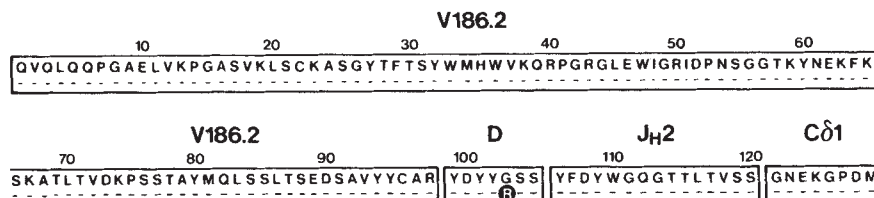


Fig. 2 Binding of $\lambda 1$ light chain to variant and wild-type heavy chain. The myeloma protein HOPC1 (IgG2a, $\lambda 1$) and IgD from B1-8.δ1 and B1-8.δ1V3 were mildly reduced with 10 mM dithioerythritol (150 min, room temperature, in the dark) and light and heavy chains separated on a Sephadex G-150 column (Pharmacia) in 1 M acetic acid with 10 mM dithioerythritol. The various fractions were dialysed first against 10 mM sodium acetate pH 5.5 and then against phosphate-buffered saline (PBS). When tested on a 7.5% SDS-polyacrylamide electrophoresis gel, the separated chains were $>95\%$ pure. HOPC1 $\lambda 1$ chains were iodinated²⁷, re-associated at 20 $\mu\text{g ml}^{-1}$ in PBS with 1% BSA at 4 °C overnight to either B1-8.δ1 (▲) or B1-8.δ1V3 (●) heavy chains (20 $\mu\text{g ml}^{-1}$) and tested for binding to an Ac38 (30 $\mu\text{g ml}^{-1}$) anti-idiotope-coated plastic plate. The $\lambda 1$ chains can bind to the anti-idiotope only when associated with a B1-8.δ1 chain. Bound $\lambda 1$ chains were counted in a γ -counter. The re-association of $\lambda 1$ to B1-8.δ1 or B1-8.δ1V3 heavy chains was inhibited by increasing amounts (abscissa) of B1-8.δ1V3 (●) or B1-8.δ1 (□) heavy chains, respectively. The two curves cross each other at roughly 50% binding, indicating a similar affinity of the two heavy chains for the $\lambda 1$ light chain.

Fig. 3 Amino-acid sequence of the B1-8.δ1V3 heavy-chain variable region. The B1-8.δ1V3 antibody was cleaved by cyanogen bromide and the three fragments of the heavy-chain variable region isolated as described for B1-8.δ1 and B1-8.δ1V3 (refs 5, 28). Because H1 (residues 1-34) and H3 (residues 82-128) are blocked by the presence of N-terminal pyroglutamic acid⁸, only H2 (residues 35-81) was sequenced directly by automated Edman degradation in a spinning-cup sequencer. Fragments H1, H2 and H3 were cleaved with trypsin or/and thermolysine and the fragments isolated by reverse-phase HPLC. The peptides were identified by amino-acid analysis and compared with the B1-8.δ1 sequence (upper line, single-letter code). The peptides covering residues 20-34 and 99-128 were sequenced in a gas-phase sequencer²⁹. The amino-acid compositions of the peptides covering residues 1-21 and 82-98 isolated from B1-8.δ1 and B1-8.δ1V3 were indistinguishable, suggesting that for these parts both heavy chains are identical. The sequence of B1-8.δ1V3 (lower, dashed line) shows at position 103 an arginine residue instead of the glycine found for B1-8.δ1. The region of the variable regions encoded by the *D* gene (*D*), *J_H2* gene (*J_H2*) and the *J_H2* border to the first constant domain (*C_H1*) are indicated by vertical lines.



Ac146, A25.9.7, Ac38, A31.90.8, A6.24, As79 and Ac106. Efficient binding was observed only in the case of NIP-BSA and antibody Ac146 (Fig. 1a). The Ac146 determinant, however, is also affected by the mutation, as the B1-8.δ1V3 antibody was about 100-fold less efficient than the wild type in inhibiting the binding of radiolabelled antibody B1-8.δ1 to antibody Ac146 (Fig. 1b). In the case of the anti-idiotopes Ac38, A25.9.7, As79 and A31.90.8, binding inhibition by the mutant was even less efficient.

Despite its drastic idiotypic alteration, the mutant antibody inhibited the binding of the B1-8.δ1 wild type to NIP-BSA with the same efficiency as the wild-type antibody itself (Fig. 1b), and the affinities of both antibodies to NIP and some related haptens were identical within the limits of the analysis (Table 1).

The mutation in B1-8.δ1V3 was localized to the heavy chain. In chain re-association experiments, mutant and wild-type heavy chains, when re-associated to λ1 chains of the myeloma HOPC1, gave rise to mutant and wild-type immunoglobulin, respectively. The mutation does not interfere with the capacity of the heavy chain to bind light chain, as the two heavy chains competed with equal efficiency for re-association to λ1 (Fig. 2). Isoelectric focusing of Fv fragments revealed a small but distinct difference between the C-terminal heavy-chain variable-region fragments (relative molecular mass 19,000) of mutant and wild-type immunoglobulin (data not shown).

The amino-acid sequence of the *V_H*, *D* and joining (*J*) segments of the B1-8.δ1V3 heavy chain, determined by the strategy developed previously for the sequencing of B1-8.δ1V1 (ref. 5), revealed a single amino-acid difference between variant and wild type; at position 103 a glycine residue is changed to an arginine residue (Fig. 3).

In the previously isolated mutant, B1-8.δ1V1 (ref. 4), recombination (conversion) between the expressed and a non-rearranged *V_H* gene had generated the mutant phenotype^{5,6}. Although we cannot exclude such an event in the case of B1-8.δ1V3, it seems more likely that in this antibody a point mutation has occurred, changing the glycine-encoding triplet GGT at position 103 (ref. 13) into CGT, encoding arginine. Point mutation is also the simplest explanation for the generation of mutants of the S107 myeloma line with changed antigen-binding specificity^{8,14}.

The observations that single amino-acid substitutions can profoundly affect idiotypic and/or antigen-binding specificity and that the *D* element has a key role in determining idiotypic specificity are not new^{3,15-20}. The main point of the present analysis is the demonstration that a single amino-acid exchange in *D* can essentially eliminate the reactivity of an antibody with the majority of the anti-idiotopes raised against it, while leaving hapten-binding specificity intact. The conservation of NIP-binding specificity in the mutant, its reactivity with the NIP group on a macromolecular carrier (BSA), and the chain re-association data argue that the mutation has left the overall structure of the antibody and the hapten-binding cleft intact and selectively changed a surface structure (arginine 103 is a surface residue according to ref. 21) of key importance for

idiotypic recognition. This structure may be directly involved in the interaction of the various anti-idiotope antibodies with the B1-8 V region. Like other protein interactions, idiotypic interactions seem in general to involve extended areas on the protein surface^{3,19}. In the present system it has been shown that, in addition to *D*, residues in *V_H* and the V region of the light chain all contribute to the reactivity of antibody B1-8 with various anti-idiotopes^{3,4,22}. It is therefore not surprising that, as far as analysed, the anti-idiotope antibodies compete with each other in binding to antibody B1-8 (ref. 11). Nevertheless, the points of contact differ for the various antibodies, as none of the target idiotopes is necessarily co-expressed with any other (refs 12, 23, 24 and our unpublished observations). Only some of the idiotopes are intimately associated with the hapten-binding site^{11,12}.

In biological terms, the B1-8.δV3 mutant demonstrates that a single amino-acid exchange can make an antibody essentially invisible for a variety of anti-idiotypic antibodies. Thus, somatic mutation can easily mediate escape from idiotypic control²⁵.

We thank S. Irlenbusch and K. Neifer for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 74) through grants to K.B. and K.R., by the Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen and by the Fazit Stiftung.

Received 18 December 1984; accepted 27 March 1985.

- McKean, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3180-3184 (1984).
- Sablitzky, F., Wildner, G. & Rajewsky, K. *EMBO J.* (in the press).
- Rajewsky, K. & Takemori, T. A. *Rev. Immun.* **1**, 569-607 (1983).
- Brüggemann, M., Radbruch, A. & Rajewsky, K. *EMBO J.* **1**, 629-634 (1982).
- Dildrop, R., Brüggemann, M., Radbruch, A., Rajewsky, K. & Beyreuther, K. *EMBO J.* **1**, 635-640 (1982).
- Krawinkel, U., Zobebein, G., Brüggemann, M., Radbruch, A. & Rajewsky, K. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4997-5001 (1983).
- Cook, W. D. & Scharff, M. D. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5687-5691 (1977).
- Cook, W. D., Rudikoff, S., Giusti, A. M. & Scharff, M. D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1240-1244 (1982).
- Reth, M., Hämmerling, G. J. & Rajewsky, K. *Eur. J. Immun.* **8**, 393-400 (1978).
- Neuberger, M. S. & Rajewsky, K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1138-1143 (1981).
- Reth, M., Imanishi-Kari, T. & Rajewsky, K. *Eur. J. Immun.* **9**, 1004-1013 (1979).
- Rajewsky, K., Takemori, T. & Reth, M. in *Monoclonal Antibodies and T Cell Hybridomas—Perspectives and Technical Advances* (eds Hämmerling, G. J., Hämmerling, U. & Kearney, J. F.) 399-409 (Elsevier, Amsterdam, 1981).
- Bothwell, A. L. M. *et al. Cell* **24**, 625-637 (1981).
- Rudikoff, S., Giusti, A., Cook, W. & Scharff, M. D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1979-1983 (1982).
- Clevinger, B., Schilling, J., Hood, L. & Davie, J. M. *J. exp. Med.* **151**, 1059-1070 (1980).
- Pollock, B. C., Kearney, J. F., Vakil, M. & Perry, R. P. *Nature* **311**, 376-379 (1984).
- Jeske, D. J., Jarvis, J., Milstein, C. & Capra, D. J. *J. Immun.* **133**, 1090-1092 (1984).
- Berek, C. *Eur. J. Immun.* **14**, 1043-1048 (1984).
- Sablitzky, F. & Rajewsky, K. *EMBO J.* **12**, 3005-3012 (1984).
- Cumano, A. & Rajewsky, K. *Eur. J. Immun.* (in the press).
- Padlan, E. A. *Q. Rev. Biophys.* **10**(1), 35-65 (1977).
- Brüggemann, M. thesis, Univ. Cologne (1984).
- Takemori, T., Tesch, H., Reth, M. & Rajewsky, K. *Eur. J. Immun.* **12**, 1040-1046 (1982).
- Rajewsky, K., Takemori, T. & Müller, C. E. in *Progress in Immunology* Vol. 5 (eds Yamamura, Y. & Tada, T.) 533-542 (Academic, New York, 1984).
- Jerne, N. K. *Annls Immun. Inst. Past., Paris* **125C**, 373-389 (1974).
- Klinman, N. R. *et al. Annls Immun. Inst. Past., Paris* **127C**, 489-504 (1976).
- Greenwood, F. C. & Hunter, W. M. *Biochem. J.* **89**, 114-117 (1963).
- Dildrop, R. & Beyreuther, K. *Nature* **292**, 61-63 (1981).
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. biol. Chem.* **256**, 7990-7997 (1981).

Ca^{2+} -dependent and Ca^{2+} -independent phagocytosis in human neutrophils

D. P. Lew*, T. Andersson†, J. Hed†, F. Di Virgilio‡, T. Pozzan‡ & O. Stendahl†

* Infectious Disease Division, Department of Medicine, University of Geneva, CH-1211 Geneva 4, Switzerland

† Department of Medical Microbiology, Linköping University Medical School, S-581 85 Linköping, Sweden

‡ CMR Unit for Study of Physiology of Mitochondria, c/o Institute of General Pathology, Via Loredan 16, 35100 Padova, Italy

The phagocytic function of neutrophils is a crucial element in host defence against invading microorganisms. Two main specific receptor-mediated mechanisms operate in the phagocyte plasma membrane, one recognizing the C3b/bi fragment of complement and the other the Fc domain of immunoglobulin G (ref. 1). There is evidence that phagocytosis mediated by these receptors differs in the number and nature of the intracellular signals generated²⁻⁵. However, the mechanism by which receptor binding is transduced into a signal that generates the formation of the phagocyte pseudopod is not known, although extensive biochemical evidence has allowed the postulate that calcium ion gradients in the peripheral cytoplasm, by interacting with calcium-sensitive contractile proteins, initiate the process of engulfment⁶. Using the high-affinity fluorescent calcium indicator quin2 both to measure and to buffer intracellular calcium ($[\text{Ca}^{2+}]_i$), we show here that in human neutrophils two mechanisms of phagocytosis coexist: a $[\text{Ca}^{2+}]_i$ -dependent and modulated phagocytosis, triggered by activation of the Fc receptor, and a $[\text{Ca}^{2+}]_i$ -independent mechanism triggered by the activation of the C3b/bi receptors.

To monitor changes in $[\text{Ca}^{2+}]_i$ during phagocytosis, human neutrophils loaded with quin2 were exposed to yeast (*Saccharomyces cerevisiae*) particles coated with IgG, C3bi or heat-inactivated serum. Figure 1 shows how $[\text{Ca}^{2+}]_i$ changes during the phagocytosis of different particles and also the kinetics of phagocytosis, expressed as a percentage of phagocytosing cells, in the same batch of quin2-loaded neutrophils. In contrast to the observations of Campbell *et al.*⁷ with latex particles, we found that with both IgG- and C3bi-coated yeasts, $[\text{Ca}^{2+}]_i$ rises after a lag time of 2–3 min, reaching a plateau at about 4–6 min and remaining elevated for >10 min. Both the lag time and the maximum rise in $[\text{Ca}^{2+}]_i$ was dependent on the yeast to cell ratio and the intracellular quin2 concentration (see below). No $[\text{Ca}^{2+}]_i$ increase was observed with unopsonized yeast that was not ingested (Fig. 1c). In several experiments, with both IgG- and C3bi-coated yeast, phagocytosis began slightly before the rise in $[\text{Ca}^{2+}]_i$ became detectable. With cells loaded with 10 μM quin2 (same conditions as for Fig. 1), the rise in $[\text{Ca}^{2+}]_i$ was detected on average within 1.8 ± 0.4 ($n=9$) for C3bi-coated yeast and 2.8 ± 0.5 min ($n=12$) for IgG-coated yeast whereas phagocytosis was already detectable at 1 min for both IgG-coated yeast ($7 \pm 3\%$ ingesting cells, $n=5$) and C3bi-coated yeast ($27 \pm 8\%$, $n=5$).

This kinetic discrepancy between the increase in $[\text{Ca}^{2+}]_i$ and phagocytosis may be explained in two ways. First, the kinetics of the $[\text{Ca}^{2+}]_i$ increase and that of yeast ingestion are similar, but because the rise in $[\text{Ca}^{2+}]_i$ occurs initially in a small proportion of cells, possibly in the cortical cytoplasm, the average value of $[\text{Ca}^{2+}]_i$ for the whole cell population, which is what quin2 actually measures, would be affected only to a small extent; only when a large number of cells becomes engaged in phagocytosis, and perhaps the increase in $[\text{Ca}^{2+}]_i$ spreads throughout the cytoplasm, is a significant average increase in $[\text{Ca}^{2+}]_i$ revealed by quin2. Furthermore, for a process like phagocytosis, which occurs as discrete events in a proportion of the whole cell population, the quin2 technique, by definition, underestimates the real rise in $[\text{Ca}^{2+}]_i$ in the cells actually engaged in particle ingestion. Figure 1 (inset) shows that soluble

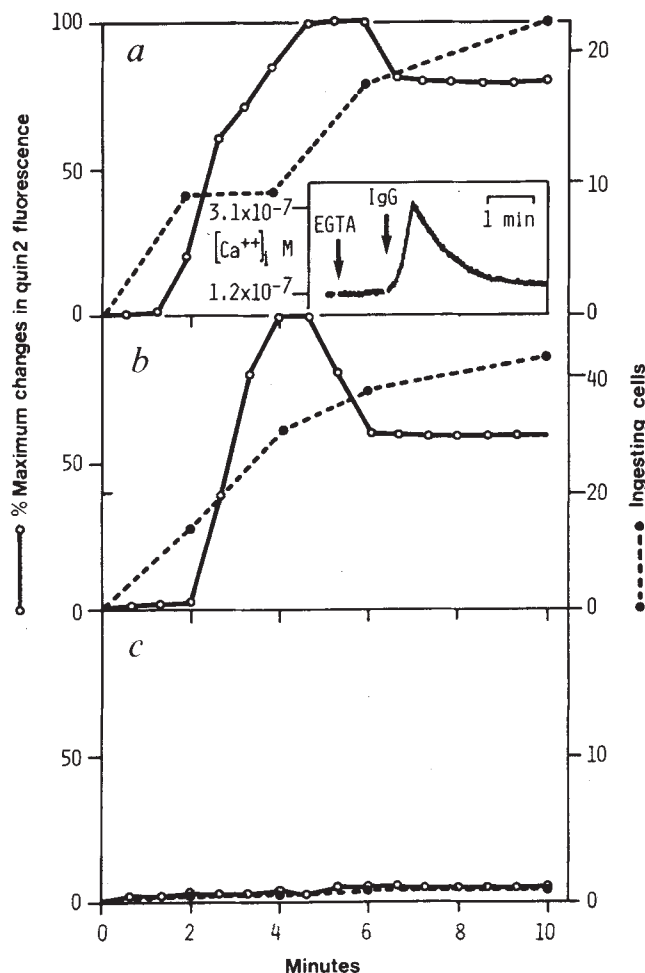
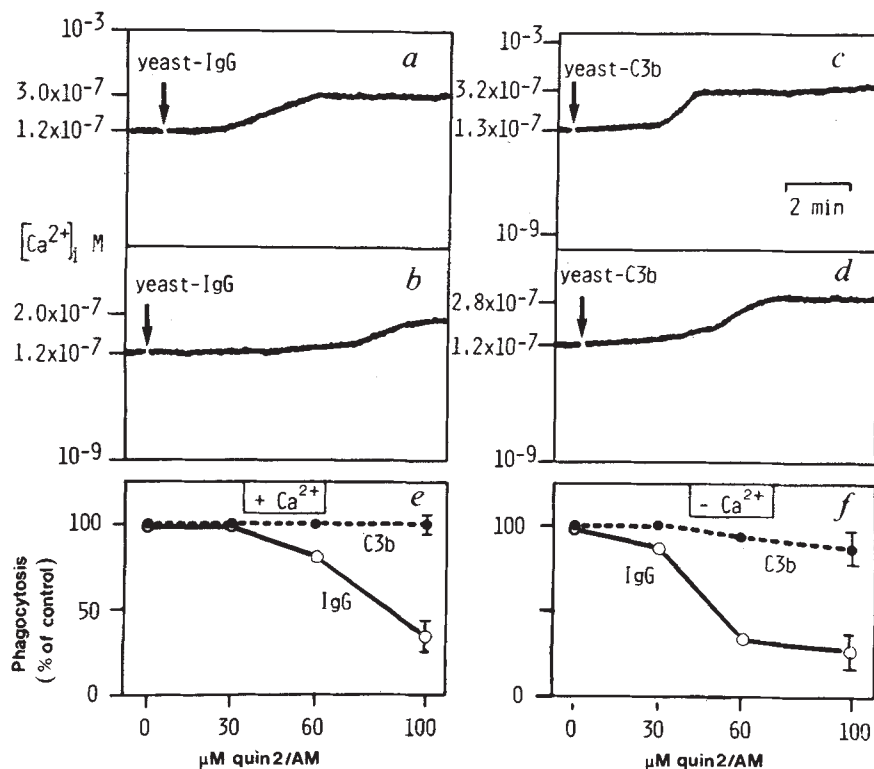


Fig. 1 Changes in $[\text{Ca}^{2+}]_i$ and phagocytosis of yeast particles by neutrophils. *a*, An experiment using yeast particles coated with rabbit anti-yeast IgG. The inset shows the effect of heat-aggregated (63 °C, 30 min) human IgG (500 $\mu\text{g ml}^{-1}$) on $[\text{Ca}^{2+}]_i$ in cells incubated in Ca^{2+} -free medium + 1 mM EGTA. *b*, *c*, Experiments using C3bi-coated yeast particles (*b*) and unopsonized yeast particles (*c*). **Methods.** Purified neutrophils were loaded with the fluorescent probe quin2/AM as described previously¹¹. Loaded cells were incubated in a spectrofluorometer (Perkin-Elmer LS3) at 37 °C with a stirring system, and $[\text{Ca}^{2+}]_i$ as well as phagocytosis in suspension were measured. Because the phagocytic assay uses fluorescein isothiocyanate (FITC)-labelled yeast particles², both experiments were performed in parallel with different aliquots from the same batch of cells. Intracellular quin2 was 0.1 nmol per 10^6 cells. Cells (5×10^6) were incubated with yeast particles (2.5×10^7) in a total volume of 2.9 ml buffer containing 138 mM NaCl, 6 mM KCl, 1 mM MgSO_4 , 1.1 mM CaCl_2 , 5 mM NaHCO_3 , 100 μM EGTA, 5.5 mM glucose, 20 mM HEPES, pH 7.4. The IgG-coated particles were opsonized by incubating yeast with purified anti-yeast IgG (20 $\mu\text{g ml}^{-1}$) in the presence of 20% heat-inactivated (56 °C, 30 min) human serum; C3bi-coated particles were opsonized by incubating the yeast with 20% fresh normal serum and unopsonized particles were incubated with 20% heat-inactivated serum. All particles were incubated at 37 °C, 30 min and washed before use. No IgG was detected on the surface of the yeast incubated with serum². Phagocytosis of FITC-labelled yeast particles was monitored by the fluorescent quenching method², where the extracellular particles are quenched by adding trypan blue. Phagocytosis is expressed as the percentage of phagocytosing cells. The uptake of C3bi-coated particles was furthermore blocked by F(ab')_2 anti-C3c (J.H. and O.S., unpublished observation). All the above experiments were performed on the same batch of quin2-loaded neutrophils. For total change in quin2 fluorescence, see Fig. 2 showing similar results.

Fig. 2 Changes in $[Ca^{2+}]_i$ and phagocytosis of C3bi- and IgG-coated yeast particles by neutrophils loaded with different concentrations of quin2. *a-d*, Neutrophils were loaded with 10 or 30 μM quin2/AM and $[Ca^{2+}]_i$ was monitored. Intracellular quin2 concentration per 10^6 cells were 0.11 and 0.34 nmol, respectively. *a, c*, Cells loaded with 10 μM ; *b, d*, cells loaded with 30 μM quin2/AM. The number of neutrophils and yeast particles were as for Fig. 1 legend. *e, f*, The effect of increasing cytosolic calcium buffering capacity by increasing intracellular quin2, on IgG- and C3bi-mediated phagocytosis. *e*, Cells incubated with various concentrations of quin2/AM in the presence of 1 mM Ca^{2+} for 60 min at 37 °C. The phagocytosis assay was performed with washed cells in the same medium containing 1 mM Ca^{2+} . *f*, Phagocytosis by cells incubated with various concentrations of quin2/AM in the absence of extracellular Ca^{2+} + 1 mM EGTA ($[Ca^{2+}]_0 = 10^{-9}$ M). This procedure depletes the cells of intracellular calcium¹². The cells were then incubated in Ca^{2+} -free medium at 37 °C for 5 min before yeast particles and Ca^{2+} (final concentration 200 μM) were added. This concentration of calcium is necessary to establish good binding of the yeast particles to the neutrophils. Phagocytosis was monitored as described in Fig. 1, and expressed as the % of phagocytosing cells after 10 min. 100% = % of phagocytosing cells in the absence of quin2. The values for IgG- and C3b-coated yeast phagocytosis in the absence of quin2 were 30% and 64%, respectively.



aggregated IgG, presumably by interacting with several receptors on most cells, initiates and mediates a rapid increase in $[Ca^{2+}]_i$ from intracellular stores, as also shown by others^{8,9}. (For further discussion of the problem of cell heterogeneity, see refs 10 and 11.) A second possibility is that phagocytosis itself does not require an increase in $[Ca^{2+}]_i$ which, if correct, would indicate that the increase in $[Ca^{2+}]_i$ observed with IgG- and C3bi-coated yeast is an event that is secondary to phagocytosis and/or receptor triggering, and not directly related to the process of particle ingestion.

To determine the role of the increase in $[Ca^{2+}]_i$ in phagocytosis, we used the fact that quin2 not only measures $[Ca^{2+}]_i$, but can also act as a high-affinity calcium buffer inside the cells. Thus, increasing the intracellular quin2 concentration, $[quin2]_i$, should decrease the amplitude and/or delay the increase in $[Ca^{2+}]_i$. The presence of increasing concentrations of quin2 inside the cells should, furthermore, counteract the formation of large $[Ca^{2+}]_i$ gradients within the cell (see also ref. 12). Thus, it can be predicted that if phagocytosis casually depends on a local, or average, rise in $[Ca^{2+}]_i$, it should be inhibited and/or delayed by increasing $[quin2]_i$. Figure 2*a-d* shows that increasing $[quin2]_i$ decreases the amplitude and delays the appearance of the yeast-induced elevation of $[Ca^{2+}]_i$. However, although the effect of quin2 loading on $[Ca^{2+}]_i$ is similar for both IgG- and C3bi-coated particles, its effect on phagocytosis is strikingly different. Figure 2*e, f* shows that phagocytosis of IgG-coated yeast is decreased markedly as a function of quin2 loading whereas ingestion of C3bi-coated particles is unaffected by increasing the intracellular calcium-buffering capacity; the percentage of ingesting cells (Fig. 2) and the number of ingested IgG-coated yeast molecules per cell are reduced to the same extent. The inhibitory effect of $[quin2]_i$ on IgG-mediated phagocytosis was more pronounced in cells that had been loaded with quin2 in the absence of extracellular Ca^{2+} ($[Ca^{2+}]_0$) + EGTA ($[Ca^{2+}]_0 = 10^{-9}$ M), a condition which decreases basal $[Ca^{2+}]_i$ and depletes intracellular Ca^{2+} stores¹². In contrast, C3bi-coated yeast particles were ingested to the same extent and with the same rate independently of the loading

conditions. In cells loaded with quin2 in Ca^{2+} -free buffer + EGTA, resting $[Ca^{2+}]_i$ was reduced to below 10 nM; the addition of Ca^{2+} (200 μM) produced a slow (5 min) increase in $[Ca^{2+}]_i$ towards basal levels, the rate of the rise being unaffected by the simultaneous addition of Ca^{2+} and opsonized yeast.

Taken together, these data suggest that Fc receptor-triggered phagocytosis depends on $[Ca^{2+}]_i$ transients, possibly localized to some specific regions of the cytoplasm, whereas C3bi receptor-mediated phagocytosis occurs at the same speed and to the same extent independently of $[Ca^{2+}]_i$. A major problem in considering C3bi-mediated phagocytosis as a completely $[Ca^{2+}]_i$ -independent process, however, is the finding that at least 100 μM Ca^{2+} must be present in the medium before yeast particle ingestion is observed. This probably does not result from an effect on $[Ca^{2+}]_i$, but reflects a role for $[Ca^{2+}]_0$ in the interaction between the surface of the cell and the yeast. It remains possible, however, that $[Ca^{2+}]_0$ is required to induce local changes in $[Ca^{2+}]_i$ or that receptor triggering sensitizes the phagocyte machinery to resting levels of $[Ca^{2+}]_i$. To exclude these two possibilities, it would be necessary to reduce $[Ca^{2+}]_i$ to 10 nM and maintain it at this low concentration throughout phagocytosis. At such a low Ca^{2+} concentration, all known Ca^{2+} -dependent processes are almost completely inhibited¹³⁻¹⁶. To study phagocytosis in the absence of extracellular calcium, C3b-opsonized endotoxin-coated oil particles were used which are efficiently ingested even at $[Ca^{2+}]_0 = 1$ nM¹⁷. Figure 3 (right panel) shows initial ingestion of C3b-opsonized oil particles by neutrophils loaded either in Ca^{2+} -containing medium ($[Ca^{2+}]_i = 120$ nM) or in Ca^{2+} -free medium + EGTA ($[Ca^{2+}]_i = 10$ nM), and tested in Ca^{2+} -free medium + EGTA. Ingestion of C3b-opsonized oil droplets, at $[Ca^{2+}]_0 = 1$ nM, was identical whether the $[Ca^{2+}]_i$ was 10 nM, where no Ca^{2+} would be released by ionophore, or $[Ca^{2+}]_i = 120$ nM and in the presence of ionophore-releasable pools. The oil particles are, unfortunately, strongly fluorescent at the quin2 wavelength, which prevents us from monitoring $[Ca^{2+}]_i$ during oil droplet phagocytosis.

The experiments shown in Fig. 3 support the $[Ca^{2+}]_i$ dependency and independency of Fc- and C3b/bi-mediated

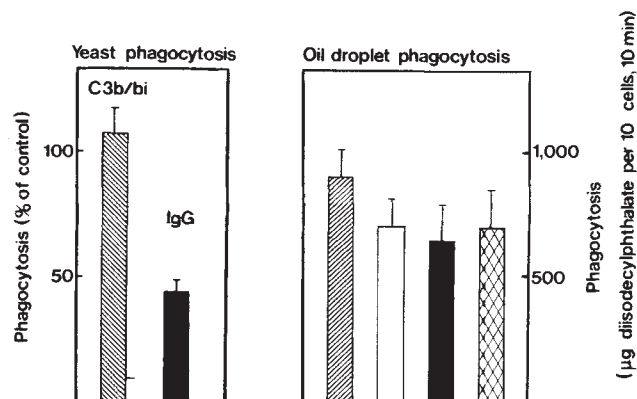


Fig. 3 *a*, Effect of the calcium ionophore ionomycin ($1 \mu\text{M}$) on phagocytosis of C3bi- or IgG-coated yeast in Ca^{2+} -containing medium. The ionophore was added 1 min before the yeast particles. Incubation time was 10 min. Phagocytosis was monitored as described in Fig. 1 legend and expressed as % of controls (=neutrophils + 0.1% dimethyl sulphoxide without ionophore). *b*, Effect of manipulation of $[\text{Ca}^{2+}]_i$ on the initial rates of ingestion of C3b-opsonized endotoxin-coated oil particles in Ca^{2+} free medium + 1 mM EGTA ($[\text{Ca}^{2+}]_0 = 10^{-9} \text{ M}$). The oil particles were prepared and opsonized as described previously¹⁸; at the end of opsonization (15 min, 37°C), 2 mM EGTA was added to the oil particles. The phagocytosis experiments were performed in the above medium containing 1 mM Mg^{2+} , 1 mM EGTA but without Ca^{2+} ($[\text{Ca}^{2+}]_0 = 10^{-9} \text{ M}$). Phagocytosis was stopped by ice-cold isotonic NaCl + 1 mM *N*-ethylmaleimide + 2 mM EDTA, and the cells were washed twice in the same buffer. $\square$, Phagocytosis in the presence of calcium by cells that were not loaded with quin2/AM. $\square$, Phagocytosis by unloaded cells in the absence of extracellular calcium (Ca^{2+} -free medium + 1 mM EGTA) ($[\text{Ca}^{2+}]_0 = 10^{-9} \text{ M}$). It was 77% of the values obtained in medium containing 1 mM Ca^{2+} , and 72% of the cells ingested oil particles. $\blacksquare$, Phagocytosis by cells loaded with quin2/AM in Ca^{2+} -containing medium and tested in Ca^{2+} -free medium + 1 mM EGTA. $\blacksquare$, Phagocytosis by quin2/AM cells, loaded and tested in Ca^{2+} -free medium + 1 mM EGTA. The intracellular quin2 concentration was 0.7 nmol per 10^6 cells. $[\text{Ca}^{2+}]_i$ in this last batch of cells was 8 nM, and no calcium could be released from intracellular stores by ionomycin. About 70% of the neutrophils contained oil droplets at the end of these four experiments after 10 min incubation.

phagocytosis, respectively. The IgG-coated yeast phagocytosis was inhibited by ~60% by preincubation of the neutrophils with the Ca^{2+} ionophore ionomycin, whereas C3bi-mediated yeast phagocytosis was completely unaffected by this treatment. Ionomycin at this concentration increases $[\text{Ca}^{2+}]_i$ to $>1 \mu\text{M}$ (not shown, see ref. 13) and should abolish any existing localized $[\text{Ca}^{2+}]_i$ transient.

Our main conclusion is that two main mechanisms of phagocytosis seem to operate in human neutrophils. First, C3b/bi-mediated phagocytosis, which occurs independently of $[\text{Ca}^{2+}]_i$, enabling phagocytosis, a motile event requiring a rearrangement of contractile proteins⁶, to occur without the involvement of calcium ions. Other signals which trigger the contractile machinery must therefore exist. Whether latex particle ingestion occurs through a similar mechanism remains to be established. The elevation in $[\text{Ca}^{2+}]_i$ induced by C3bi-coated yeast should thus be considered a secondary event probably not directly related to phagocytosis. Second, Fc receptor-mediated phagocytosis, which possibly requires localized changes in $[\text{Ca}^{2+}]_i$ and depends on the presence of this ion at least at physiological concentrations in the cytosol. The reduction, but not complete inhibition, of IgG-mediated phagocytosis by high quin2 loading or ionomycin in Ca^{2+} -containing medium suggests that Fc-mediated phagocytosis may occur, at least in part, through a mechanism that does not require major changes in $[\text{Ca}^{2+}]_i$. However, we cannot eliminate the possibility that discrete local changes in $[\text{Ca}^{2+}]_i$ close to the plasma membrane are

sufficient to trigger some ingestion, or that some IgG-independent ingestion of the yeast particles may occur.

We thank Ms Antoinette Monod, Kerstin Falk and Anita Lönn for technical assistance. This work was supported by grants from the NSF (3.990.0.84), the Swedish MRC (5968), King Gustav V Foundation, the Swedish Society of Medicine and the Italian Ministry of Public Education. P.D.L. is a recipient of a Max Cloetta Career Development Award and T.A. of a postdoctoral fellowship from the Swedish MRC.

Received 18 December 1984; accepted 12 March 1985.

1. Silverstein, S. C., Steinman, R. M. & Cohn, Z. A. *A. Rev. Biochem.* **46**, 665-722 (1977).
2. Hed, J. & Stendahl, O. *Immunology* **45**, 727-736 (1982).
3. Newman, S. L. & Johnston, R. B. Jr *J. exp. Med.* **123**, 1839-1846 (1979).
4. Wright, S. D. & Silverstein, S. C. *J. exp. Med.* **158**, 2016-2023 (1983).
5. Yamamoto, K. & Johnston, R. B. Jr *J. exp. Med.* **159**, 406-416 (1984).
6. Southwick, F. S. & Stossel, T. P. *Semin. Hemat.* **20**, 305-321 (1983).
7. Campbell, A. K. & Hallett, M. B. *J. Physiol., Lond.* **358**, 537-550 (1983).
8. Korchack, H. M. *et al. J. Biol. Chem.* **259**, 4076-4080 (1984).
9. Young, J. D.-E., Ko, S. S. & Cohn, Z. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5430-5434 (1984).
10. Tsien, R. Y., Pozzan, T. & Rink, T. J. *Nature* **295**, 68-71 (1982).
11. Tsien, R. Y., Pozzan, T. & Rink, T. J. *J. Cell Biol.* **94**, 325-334 (1982).
12. Di Virgilio, F., Lew, P. D. & Pozzan, T. *Nature* **310**, 691-693 (1984).
13. Lew, P. D., Wollheim, C. B., Waltvogel, F. A. & Pozzan, T. *J. Cell Biol.* **99**, 1212-1220 (1984).
14. Yin, H. L. & Stossel, T. P. *Nature* **281**, 583-585 (1979).
15. Cheung, W. Y. *Science* **207**, 19-27 (1980).
16. Niggli, V., Adunyah, E. S. & Carafoli, E. *J. Biol. Chem.* **256**, 8588-8592 (1981).
17. Stossel, T. P. *J. Cell Biol.* **58**, 346-356 (1973).
18. Stossel, T. P., Alper, C. A. & Rosen, F. S. *J. exp. Med.* **137**, 690-705 (1973).

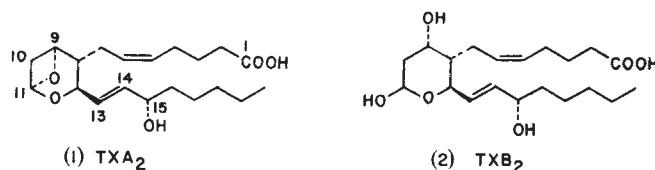
Synthesis and structure of the platelet aggregation factor thromboxane A_2

Shripad S. Bhagwat, Philip R. Hamann, W. Clark Still*, Stuart Bunting & Frank A. Fitzpatrick†

* Department of Chemistry, Columbia University, New York, New York 10027, USA

† Lipids Research, The Upjohn Company, Kalamazoo, Michigan 49001, USA

In 1975, Hamberg *et al.*¹ reported evidence for the existence of an unstable platelet-aggregating factor which they named thromboxane A_2 (TXA_2) and for which they proposed a novel bicyclic oxetane structure (1, below) based on the short half-life of the factor ($t_{1/2}(37^\circ\text{C}) = 32 \text{ s}$ at pH 7.4) and the isolation of degradation products related to thromboxane (TXB_2) (2, below). As natural TXA_2 has not yet been isolated and characterized as a pure compound, we have synthesized the proposed structure (1) from TXB_2 and compared its biological properties with those of authentic, biologically generated material. Here we present evidence that synthetic material having structure (1) is indistinguishable from platelet-derived TXA_2 in various biological assays and that the proposed structure (1) for TXA_2 is correct.



The acid-lability of TXA_2 is believed to be a result of the highly strained nature of the bicyclic oxetane acetal nucleus and it was only recently that the hetero-unsubstituted ring system, a 2,6-dioxabicyclo[3.1.1]heptane, was isolated². (References 3-6 describe the preparation of heterosubstituted derivatives.) Bhagwat *et al.*² prepared the TXA_2 model (3) by a modified Mitsunobu⁷ etherification $[(\text{MeO})_3\text{P}, \text{EtO}_2\text{CN}=\text{NCO}_2\text{Et}, \text{CH}_2\text{Cl}_2]$ of compound (4), followed by mild free-radical reduction (Bu_3SnH , C_6D_6 , sun-lamp irradiation) and, auspiciously, (3)

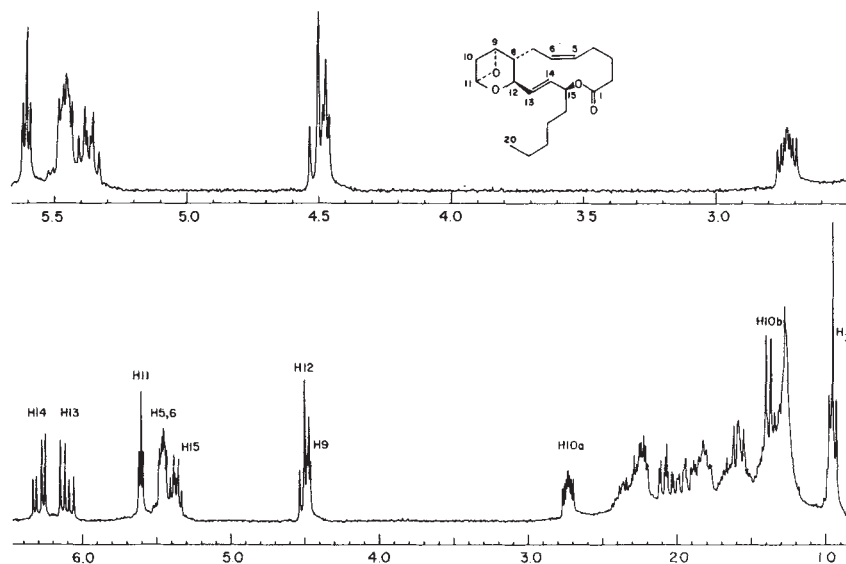


Fig. 1 270-MHz ^1H -NMR spectrum of synthetic 1,15-anhydro-TXA₂ in benzene-D₆. The lower trace shows the complete spectrum, with assignments being confirmed by double-resonance decoupling studies. The upper trace shows the expanded spectrum segment spanning 2.5–5.7 p.p.m.

was found to undergo hydrolysis in the Krebs medium⁸ at a rate comparable with that reported for TXA₂.

To prepare compound (1) by an analogous procedure, problematic C₁₀ free-radical cyclization onto the C₁₃, C₁₄ olefin in the penultimate precursor of (1) had to be avoided. This undesired reaction pathway could be eliminated by forming a conformationally restricted 1,15-macrolactone which prevented approach of the C₁₃, C₁₄ olefinic linkage to the C₁₀ free-radical intermediate in the debromination reaction. Thus, TXB₂ was converted to the 1,15-macrolactone (5) using the Corey-Nicolaou procedure (refs 9, 10 and see also refs 11–13), then dehydrated to the β -hydroxy enol ether (6) using 2-chloro-1-methylpyridinium iodide¹⁴. Formation of bromohydrin (*N*-bromosuccinimide, 1% NaHCO₃, H₂O/Et₂O; 51% yield overall from TXB₂) and oxetane [(MeO)₃P, EtO₂CN=NC₂Et, CH₂Cl₂; 20% yield] then yielded 1,15-anhydro-10-bromo-TXA₂ (7, X=Br). Finally, polymer-bound tin hydride¹⁵ reduction and filtration of the tin reagent gave the crystalline, acid-labile 1,15-anhydro-TXA₂ structure (7, X=H, melting point (pentane) = 81–82 °C) in high yield. Compound (7) (X=H) displayed the characteristic bicyclic oxetane ^1H - and ^{13}C -NMR resonances which closely matched those of compound (3) (the 270-MHz ^1H -NMR spectrum is shown in Fig. 1). The novel bicyclic oxetane structure of compound (7) was confirmed by single-crystal X-ray crystallography.

Although macrolide (7, X=H) did not show the biological properties of TXA₂, it was saponified in NaOD/D₂O/CD₃OD (30 min, 22 °C) as evidenced by ^1H -NMR, which showed opening of the 1,15-macrolactone with retention of the bicyclic oxetane nucleus, to yield the synthetic compound (1) (TXA₂S) as its sodium salt, which did show the biological properties of

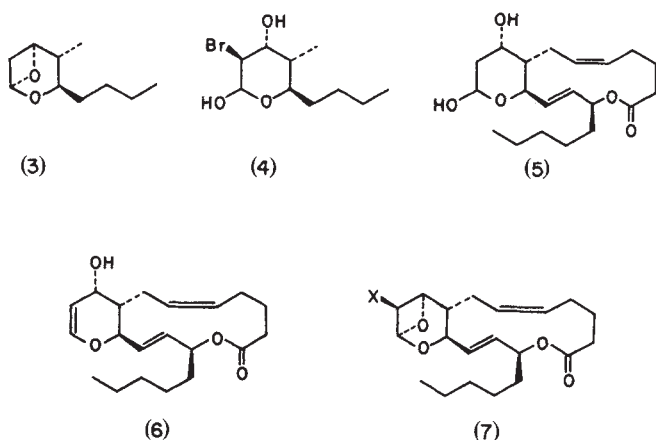
TXA₂. TXA₂S was also prepared by saponification of compound (7) (X=H) with Me₃SiOK¹⁶ in tetrahydrofuran and isolated as an amorphous solid by solvent evaporation. Although TXA₂S underwent rapid decomposition in neutral or acidic solutions, it was relatively stable in the basic saponification solutions described and could be stored for up to 1 week at –20 °C without significant loss of activity.

TXA₂S at concentrations of 16–100 ng ml^{–1} caused a concentration-dependent aggregation of human platelet-rich plasma (PRP). Irreversible aggregation occurred with 50–100 ng TXA₂S per ml for all cases examined. TXA₂S was found to be ~10 times more potent than the prostaglandin endoperoxide PGH₂ at inducing aggregation of human PRP. These comparative potencies (shown in Fig. 2) agree with published data^{17,18}. TXA₂S (10 $\mu\text{g ml}^{-1}$) decayed in human plasma with first-order kinetics, with a half-life of 3.9 min at 37 °C, consistent with values of 3 min¹⁹ and 5–6 min²⁰ determined by indirect methods for biologically generated material. The aggregatory effects of TXA₂S were confined to platelets. At 2 $\mu\text{g ml}^{-1}$ it did not aggregate human neutrophils nor did it stimulate formation of leukotriene B₄.

A selective thromboxane synthetase inhibitor with no accompanying receptor-level antagonist properties (9,11-iminoepoxy-prosta-5,13-dienoic acid, U-54701) at 5–15 $\mu\text{g ml}^{-1}$ had no effect on platelet aggregation induced by TXA₂S. In contrast, a related compound with receptor-level antagonist properties (9,11-diazoprosta-5,13-dienoic acid, U-51605) at 1–3 $\mu\text{g ml}^{-1}$ inhibited TXA₂-induced aggregation in a concentration-dependent manner²¹.

Hydrolysis of TXA₂S in 0.1 M phosphate buffer, pH 7.4, eliminated its biological activity and produced a substance that was immunologically indistinguishable from synthetic TXB₂. Radioimmunoassay determinations of hydrolysed TXA₂S yielded values of 84 ± 21 , 60 ± 16 , 39 ± 11 and 25 ± 8 pg TXB₂ (mean $\pm$ s.d. at 95% confidence level, $n=3$) which correspond to the predicted values of 125, 62, 31 and 15 pg, respectively. The half-life of TXA₂S at 37 °C in Krebs solution, measured by bioassay on vascular tissue, was 30 s, which corresponds closely to previously reported half-lives for biologically generated material^{1,17}.

The substance that induces contraction of rabbit aorta, detected by Piper and Vane²² in the effluent of perfused guinea pig lung during anaphylaxis, was shown to be a mixture of TXA₂ and prostaglandin endoperoxides. Similarly, TXA₂ generated by platelets or a partially purified platelet enzyme, has potent vascular contracting activity. Thus, both natural and synthetic TXA₂ contracted all vascular tissue tested including Krebs-superfused²³ rabbit aorta, pulmonary, mesenteric and coeliac arteries and bovine coronary arteries. On most tissues,



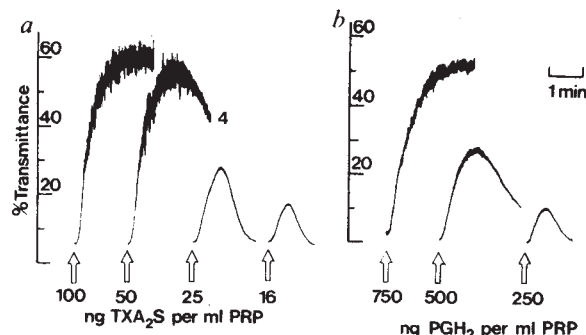


Fig. 2 Dose-response curves for aggregation of platelets in human platelet-rich plasma (PRP) for synthetic (a) and biologically generated (b) TXA_2 . Whole blood (45 ml) was collected over 3.8% w/v trisodium citrate (5 ml). PRP (20 ml) was isolated from blood after centrifugation at 200g for 10 min at 25°C. PRP suspensions (1.0 ml) were incubated at 37°C for 2 min, then transferred to an aggregometer cuvette containing TXA_2S (10–200 ng) or prostaglandin endoperoxide PGH_2 (100–1,000 ng). Platelet suspensions were stirred at 1,100 r.p.m. at 37°C and aggregation was monitored as described by Born²⁵. To verify that the effects were attributable solely to exogenous TXA_2 or PGH_2 , and to exclude any contribution to aggregation from activation of endogenous arachidonic acid metabolism, PRP containing 250 μM aspirin was used.

2–3 ng of TXA_2S was the threshold dose for contraction, generating ~200 mg of tension, with higher doses generating more than 2 g of tension in rabbit aorta and 4 g of tension in bovine coronary arteries. The potency of TXA_2S compared well with the activity of biologically produced TXA_2 , calculated from the amount of PGH_2 originally in the incubation (assuming 100% conversion). Rabbit mesenteric and coeliac arteries respond to PGH_2 with a small, short-lived contraction, followed by larger, longer-lasting relaxation²⁴, whereas TXA_2S and biologically generated TXA_2 gave only constriction.

Although these experiments do not exclude the formal possibility that compound (1) is not TXA_2 itself but is converted *in vivo* to a different material which is the endogenous TXA_2 , the qualitative and quantitative activity and the half-life of TXA_2S provide strong evidence that the Samuelsson¹ oxetane structure (1) for TXA_2 is correct.

The synthetic work described here was carried out by S.S.B., P.R.H. and W.C.S. at Columbia University. These authors thank Dr John Pike of the Upjohn Company for a generous supply of TXB_2 and the NIH for support (HL25634). Biological studies were carried out by S.B. and F.A.F. at the Upjohn Company.

Received 15 February; accepted 22 March 1985.

Tissue localization and chromosomal assignment of a serum protein that tracks the cystic fibrosis gene

Veronica van Heyningen, Caroline Hayward*, Judy Fletcher & Christine McAuley*

MRC Clinical and Population Cytogenetics Unit and *Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK

The basic gene defect in the autosomal recessive disorder cystic fibrosis has not been identified, and no firm linkage of the disorder to any other marker has been reported. However, a serum protein abnormality present in unaffected heterozygotes as well as in affected homozygotes has been described¹, and immunological quantitation of this protein, termed cystic fibrosis antigen, allows the three genotypes to be distinguished^{2,3}. We show here that an immunologically indistinguishable protein is present at high concentrations in granulocytes from normal and cystic fibrosis individuals as well as in myeloid leukaemia cells. Somatic cell hybrids between the mouse myeloid stem-cell line WEHI-TG and myeloid leukaemia cells express cystic fibrosis antigen only when human chromosome 1 is present.

Cystic fibrosis is one of the most common serious genetic diseases in Caucasian populations: 1 in 2,000 live births produces an affected child. Assuming that homozygosity for mutant allele(s) at a single locus gives rise to affected individuals⁴, the expected heterozygote frequency in these populations is about 1 in 22. The major abnormalities are those of exocrine secretion⁵. Raised sweat sodium and chloride concentration is currently the best diagnostic test for the disease⁶, but clinical problems arise because of chronic obstructive lung disease and exocrine pancreatic insufficiency leading to malabsorption and neonatal intestinal blockage⁵. Despite much effort to define the basic gene defect, no obvious candidate has been found among the many, mainly secondary, observed metabolic abnormalities. Therefore, the study of any abnormality which is manifested in a gene dosage-dependent manner in the symptomless obligate heterozygote as well as in the affected homozygote may lead to the identification of a primary defect and hence to the aberrant gene in cystic fibrosis⁷.

We have followed a serum protein first described by Wilson *et al.*¹ as a doublet band at pI 8.4 on isoelectric focusing (IEF) in the presence of urea. This doublet band is present in serum from cystic fibrosis homozygotes and heterozygotes but not in normal serum, a finding confirmed by several other groups^{8–10}. Manson and Brock² produced a guinea pig antiserum to the doublet band cut out of IEF gels. When used in rocket immunoelectrophoresis, this antiserum distinguishes all three genotypes: homozygotes have high, heterozygotes intermediate and normals very low (or absent) concentrations of this protein, cystic fibrosis antigen. But the exact method of serum collection is critical for accurate quantitation of the antigen and hence for heterozygote detection. The performance of the rocket assay on optimally collected serum samples is described by Bullock *et al.*³. Blood from normal individuals collected in the presence of the serine protease inhibitor benzamidine exhibits high levels of cystic fibrosis antigen. Heparin- or EDTA-collected plasma samples, on the other hand, possess antigen levels expected on the basis of genotype, but produce poor rockets.

Attempts have been made to purify the antigen to identify it biochemically and to produce monoclonal antibodies which would improve and simplify the heterozygote detection technique, but with increasing purification the antigen was repeatedly lost. Partial purification by gel-permeation on Sephacryl S-300, however, is possible. The antigen can be recovered in fractions strongly contaminated with albumin, suggesting that it has a relative molecular mass of ~70,000. To attempt to eliminate this contamination, several probable tissue and cell

- Hamberg, M., Svensson, J. & Samuelsson, B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2994–2998 (1975).
- Bhagwat, S. S., Hamann, P. R. & Still, W. C. *Tetrahedron Lett.* **26**, 1955–1958 (1985).
- Ito, H., Eby, R., Kramer, S. & Schuerch, C. *Carbohydr. Res.* **86**, 193–202 (1980).
- Varma, A. J. & Schuerch, C. *J. org. Chem.* **46**, 799–803 (1981).
- Kong, F. & Schuerch, C. *Carbohydr. Res.* **112**, 141–147 (1983).
- Fried, J., Hallinan, E. A. & Szeto, M. J. *J. Am. chem. Soc.* **106**, 3871–3872 (1984).
- Mitsunobu, O. *Synthesis*, 1–28 (1981).
- Krebs, H. A. & Henseleit, K. *Hoppe-Seyler's Z. physiol. Chem.* **210**, 33 (1932).
- Corey, E. J. & Nicolaou, K. C. *J. Am. chem. Soc.* **96**, 5614–5616 (1974).
- Corey, E. J. & Clark, D. A. *Tetrahedron Lett.* 2875–2878 (1979).
- Corey, E. J., Nicolaou, K. C. & Melvin, L. S. *J. Am. chem. Soc.* **97**, 653–654 (1975).
- Narasaka, K., Maruyama, K. & Mukaiyama, T. *Chem. Lett.*, 885–888 (1978).
- Bundy, G. L. *et al. J. med. Chem.* **26**, 1089–1099 (1983).
- Mukaiyama, T., Usui, M. & Shimada, E. *Chem. Lett.*, 1045–1048 (1975).
- Weinshenker, N. M., Crosby, G. A. & Wong, J. Y. *J. org. Chem.* **40**, 1966–1971 (1975).
- Laganis, E. D. & Chenard, B. L. *Tetrahedron Lett.* **25**, 5831–5834 (1984).
- Needleman, P. *et al. Nature* **261**, 558–560 (1976).
- Needleman, P., Minkes, M. & Raz, A. *Science* **193**, 163–165 (1976).
- Folco, G., Granstrom, E. & Kindahl, H. *FEBS Lett.* **82**, 321–324 (1977).
- Smith, J. B., Ingeman, C. & Silver, M. *J. clin. Invest.* **58**, 1119–1122 (1976).
- Fitzpatrick, F. *et al. Biochim. biophys. Acta* **573**, 238–244 (1979).
- Piper, P. J. & Vane, J. R. *Nature* **223**, 29–35 (1969).
- Vane, J. R. *Br. J. Pharmacol.* **12**, 344–349 (1957).
- Bunting, S., Moncada, S. & Vane, J. R. *Br. J. Pharmacol.* **57**, 462 (1976).
- Born, G. V. R. *Nature* **194**, 927–929 (1962).

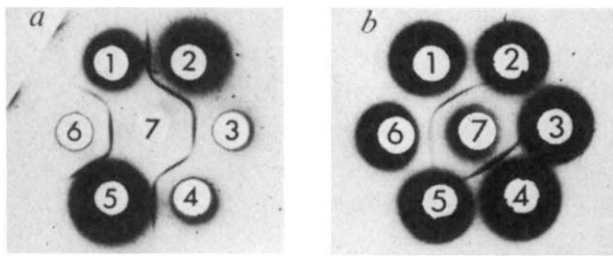


Fig. 1 Ouchterlony double-diffusion analysis for the presence of cystic fibrosis antigen. *a*, 1, Pooled sera from 15 cystic fibrosis homozygotes; 2, guinea pig antiserum raised against material cut out of the cystic fibrosis-specific doublet band region of IEF gels¹ (GP anti-IEF); 3, guinea pig antiserum raised against partially purified antigen from myeloid leukaemia cells (GP anti-CML); 4, antiserum raised as in 3, but in rabbit; 5, citrated plasma from a cystic fibrosis patient; 6, GP anti-CML; 7, freeze-thawed, spun lysate in Dulbecco A buffer of peripheral leukocytes (at 10^6 cells per ml) from a CML patient who was the human parent in the WEMAD hybrids. *b*, 1-5, Cell lysates prepared as for the CML cells above. 1, WEMAD 1.1.9.1; 2, WEMAD 1.1.6.3; 3, NYX 3.1 (see text); 4, WEH 3127.6.1; 5, WEP 1342.2.3; 6, cystic fibrosis serum pool; 7, GP anti-IEF. For assessment of antigen status see Fig. 2. Wells contained $10\ \mu\text{l}$ of appropriately diluted antiserum or antigen. Diffusion was carried out for 24 h at room temperature in 1.5% Difco Noble agar in Dulbecco A buffer containing 0.02% azide; the gels were washed extensively in buffer with 0.05% Tween 20, dried down on Gelbond and stained with Coomassie Blue.

types were examined as potential antigen sources⁵ by rocket immunoelectrophoresis and by immunodiffusion. Peripheral leukocytes from chronic myeloid leukaemia (CML) patients proved consistently to be strongly positive for cystic fibrosis antigen, and antigen from normal granulocytes was immunologically identical. It remains to be seen whether the presence of antigen in tissue lysates from adult liver, kidney and brain is caused by blood contamination of these tissues or whether it is intrinsic to tissue-specific cells. Lymphoblastoid cell lines from normal and cystic fibrosis individuals are antigen negative. Ouchterlony double diffusion demonstrates the immunological identity of cystic fibrosis antigen from serum and CML cells (Fig. 1*a*), and also shows that antiserum raised in rabbit and guinea pig against partially purified antigen from CML lysates detects the antigen in appropriate serum samples.

It has been shown previously that at least some human granulocyte-specific products are expressed in somatic cell hybrids between the mouse myeloid stem-cell line WEHI-TG and CML cells¹¹. To observe segregation of cystic fibrosis antigen with a human chromosome, we produced a series of independent hybrid lines between WEHI-TG and cells from four different myeloid leukaemia patients. As soon as sufficient cells were available from each hybrid line, cell lysates were tested by Ouchterlony assay for the presence of antigen. (WEHI-TG and mature mouse granulocytes are negative in this assay). Negative and positive hybrids were subcloned repeatedly to achieve reasonable chromosomal stability. Subclones were assayed for cystic fibrosis antigen and analysed for the presence of human chromosomes within the space of one to three transfer generations. Figure 1*b* shows the clear-cut results seen with these hybrid lines as long as the equivalence point between each cell lysate and the guinea pig antiserum was predetermined by previous titration. The results of the complete segregation analysis on 15 subclones from 4 independent hybrid lines are summarized in Fig. 2. Antigen segregates with human chromosome 1 with complete concordance. No other chromosome is consistently present with this antigen, suggesting that the structural gene for this protein is on chromosome 1. Note that NYX 3.1 (a mouse myeloma $\times$ human B-lymphoblastoid line somatic-cell hybrid containing chromosome 1 at high frequency) does not express the antigen. This observation agrees with the notion that this serum protein is a differentiated cell product.

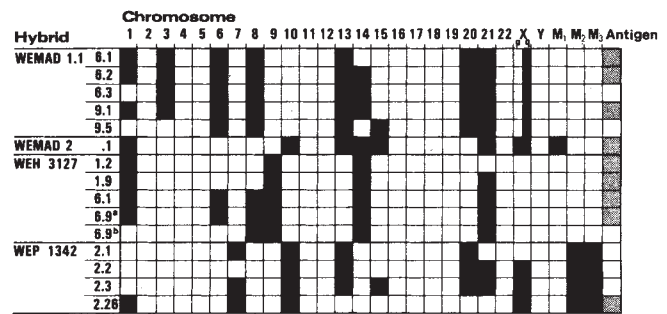


Fig. 2 The segregation of human chromosomes and cystic fibrosis antigen in subclones from a series of four independent hybrid lines between the mouse myeloid stem cell line WEHI-TG (ref. 11) and human myeloid leukaemia cells. Filled squares denote the presence of human chromosome in $>30\%$ of karyotyped cells (>16 cells for each clone). Open squares denote the absence of the chromosome in $>95\%$ of cells. Each series of marked subclones (WEMAD 1, WEMAD 2, WEH and WEP) is derived from a hybrid line which arose from an independent fusion event. Fusion was effected between leukaemic cells and WEHI-TG (at 10:1 ratio) with 0.5 ml 25% polyethylene glycol 1500 containing 1,250 haemagglutinating units of β -propiolactone-inactivated Sendai virus. Hybrid cells were selected for by selection using hypoxanthine-aminopterin-thymidine¹³. Independent hybrids arose in separate wells. Subcloning was by limiting dilution and carried out repeatedly until reasonable chromosomal stability was established. Subclones were then grown up so that cells within a few transfer generations were preserved in liquid nitrogen, pelleted for antigen assay and karyotyped by sequential trypsin banding¹⁴ and Giemsa-11 staining¹⁵. The subclone WEH 3127.6.9 contained cystic fibrosis antigen and chromosome 1 was present in 32% of cells on first testing (*a*). Six transfer generations later (*b*) both the antigen and chromosome 1 were absent from WEH 3127.6.9. Unidentifiable human marker chromosomes M₁, M₂ and M₃ were found in some hybrids. None seemed to be derived from chromosome 1.

As with other autosomal recessive diseases, the defect in cystic fibrosis presumably results from the absence of a normal biological function which arises only if both copies of a gene are aberrant. Assuming that the accumulation of cystic fibrosis antigen is closely related to the basic genetic defect, what sort of mutation could account for the observed pattern: high, intermediate and zero serum levels, in cystic fibrosis homozygotes, heterozygotes and normals respectively, of a protein immunologically indistinguishable (with available antisera) from a normal granulocyte product?

Increased granulocyte fragility which might result from a membrane defect is an unlikely explanation because heterozygotes have increased antigen levels and yet are symptom-free. Suggestions that homozygote and heterozygote sera contain factor(s) which induce increased leukocyte degranulation¹² are not supported by our observation that serum levels of lactoferrin (a specific granule protein) are unaffected by cystic fibrosis gene status (D. Brock, S. Chung and V. van H., unpublished). Another possibility is that a protein that is routinely released into the circulation from granulocytes needs to be processed (for example, by proteolytic cleavage) to fulfil a normal biological function and in the course of this it becomes antigenically unrecognizable. The gene dose dependence of antigen level again makes it unlikely that the mutation resides in a processing enzyme as this would require the aberrant step in the processing sequence to be rate-limiting. The simplest explanation is that the accumulated protein itself is the product of the cystic fibrosis gene and it is altered so that it cannot be normally processed.

The probable site for the mutation, then, is a region of polypeptide chain which acts as a site for a specific proteolytic cleavage step, probably one of several sequential steps involved in releasing a biologically active product from a precursor. This scheme would explain the presence of the high level of antigen and the absence of an essential function in affected homozygotes. Heterozygotes, on the other hand, would possess one normally

processed copy of the cystic fibrosis gene and one mutant copy which would give rise to a half dose of the antigen.

Resolving the biochemical identity of the cystic fibrosis antigen should provide insight into the nature of the disease and allow critical assessment of the genetic hypothesis described here. We are attempting the immunological purification of the antigen, using low-affinity monoclonal antibodies, and are also screening for high-affinity antibodies which will permit the development of a more reliable heterozygote detection assay. Attempts are also in progress to isolate the cystic fibrosis antigen gene by complementary DNA cloning.

We thank Dr David Brock for encouragement and helpful discussion, Dr A. Geurts van Kessel for the gift of WEHI-TG cells, and Drs A. E. Dewar, C. A. Parkin, N. C. Allan, A. Birch and J. A. Raeburn for supplying blood samples. This work was supported in part by the Cystic Fibrosis Research Trust.

Received 15 February; accepted 2 April 1985.

1. Wilson, G. B., Fudenberg, H. H. & Jahn, T. L. *Pediatr. Res.* **9**, 635-640 (1975).
2. Manson, J. C. & Brock, D. J. H. *Lancet* **i**, 330-331 (1980).
3. Bullock, S., Hayward, C., Manson, J. C., Brock, D. J. H. & Raeburn, J. A. *Clin. Genet.* **21**, 336-341 (1982).
4. Romeo, G. in *Cystic Fibrosis: Horizons* (ed. Lawson, D.) 155-164 (Wiley, Chichester, 1984).
5. Talamo, R. C., Rosenstein, B. J. & Berninger, R. W. in *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B., Wyngaarden, J. B., Frederickson, D. S., Goldstein, J. L. & Brown, M. S.) 1889-1917 (McGraw-Hill, New York, 1983).
6. Denning, C. R. *et al. Pediatrics* **66**, 752-757 (1980).
7. van Heyningen V. *Nature* **311**, 104-105 (1984).
8. Nevin, G. B., Nevin, N. C., Redmond, A. O., Young, I. R. & Tully, G. W. *Hum. Genet.* **56**, 387-389 (1981).
9. Brock, D. J. H., Hayward, C. & Super, M. *Hum. Genet.* **60**, 30-31 (1982).
10. Grataroli, R., Guy-Crotte, O., Galabert, C. & Figarella, C. *Pediatr. Res.* **18**, 130-133 (1984).
11. Geurts van Kessel, A. H. M., Tetteroo, P. A. T., von dem Borne, A. E. G., Hagemeijer, A. & Bootsma, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3748-3752 (1983).
12. Conod, E. J., Conover, J. H. & Hirschhorn, K. *Pediatr. Res.* **9**, 724-729 (1975).
13. Littlefield, J. W. *Science* **145**, 709-710 (1964).
14. Gallimore, P. H. & Richardson, C. R. *Chromosoma* **41**, 259-263 (1973).
15. Bobrow, M. & Cross, J. *Nature* **251**, 77-79 (1971).

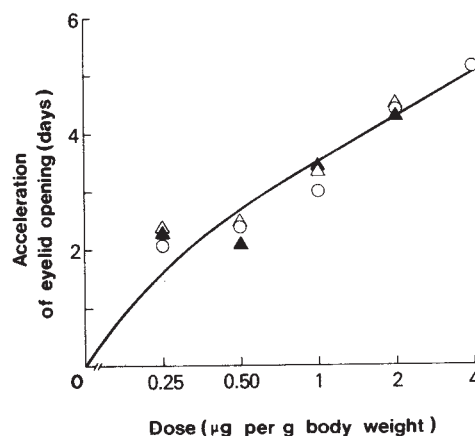


Fig. 1 Dose-response curve for activity of human TGF- α (▲), human EGF (△) and mouse EGF (○) in accelerating eyelid opening in newborn Swiss mice.

Human transforming growth factor- α causes precocious eyelid opening in newborn mice

Joseph M. Smith*, Michael B. Sporn*,
Anita B. Roberts*, Rik Derynck†,
Marjorie E. Winkler† & Harry Gregory‡

* National Cancer Institute, NIH, Bethesda, Maryland 20205, USA

† Genentech, Inc., South San Francisco, California 94080, USA

‡ ICI Ltd, Pharmaceuticals Division, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Both murine and human epidermal growth factors (EGFs) are known to cause precocious opening of the eyelids in newborn mice¹⁻³. Another set of peptides that are structurally and functionally homologous to murine and human EGFs are the murine and human type- α transforming growth factors (TGF- α s)⁴⁻⁷. TGF- α s have been found in many cancer cells and it has been suggested that their autocrine action may play an important part in malignant transformation⁸⁻¹¹. In several *in vitro* systems murine and human TGF- α s are functionally interchangeable with murine and human EGFs^{6,12-14}. However, the *in vivo* activity of the TGF- α s has not been characterized, as only small amounts of these peptides were available until recently. The cloning of the gene for human TGF- α and its expression in *Escherichia coli*⁶ now allow us to demonstrate that human TGF- α is as active as murine EGF in promoting eyelid opening in newborn mice. Furthermore, we show in a dose-dependent eyelid opening assay that human EGF is as potent as its murine homologue with respect to this biological property.

The original discovery and purification of murine EGF¹ was based on the measurement of the intrinsic ability of this molecule

to accelerate eyelid opening as its primary activity. This property formed the basis for the discovery of the identity of human urogastrone and human EGF and their functional as well as structural homology to murine EGF^{2,3}. TGF- α s were first described in retrovirus-transformed cells⁸ and have subsequently been found to be secreted by a wide variety of rodent and human tumour cells, suggesting that aberrant expression of these peptides has a role in transformation. It has been speculated that TGF- α functions as an embryonic form of EGF, but it is uncertain what physiological role, if any, TGF- α has in post-embryonic life. Both rat and human TGF- α s have recently been isolated as pure substances and their amino-acid sequences determined⁴⁻⁷; they have significant homologies with mouse and human EGFs, particularly with regard to the location of their six cysteine residues, which are important determinants of molecular shape through the formation of disulphide bridges. Human and murine EGF and TGF- α all bind to a common receptor, generally referred to as the EGF receptor, and the qualitative and quantitative activity of all four peptides in many *in vitro* assays is essentially the same^{6,13-16}.

In the experiments reported here, 315 newborn mice were injected with human TGF- α , human EGF, mouse EGF or the vehicle for the growth factors (phosphate-buffered isotonic saline, PBS) as a control. Each of the three growth factors was shown to be pure by its unique amino-acid composition, particularly by the absence of specific amino acids. Human TGF- α

Table 1 Acceleration of eyelid opening by TGF- α and EGF

Dose (μ g)	Mouse EGF	Human EGF	Human TGF- α
4.0	8.8 $\pm$ 0.2 (4)		
2.0	9.6 $\pm$ 0.1 (35)	9.4 $\pm$ 0.1 (21)	9.6 $\pm$ 0.3 (18)
1.0	10.9 $\pm$ 0.1 (20)	10.6 $\pm$ 0.1 (18)	10.5 $\pm$ 0.2 (21)
0.50	11.5 $\pm$ 0.1 (16)	11.4 $\pm$ 0.1 (18)	11.8 $\pm$ 0.2 (18)
0.25	11.8 $\pm$ 0.2 (12)	11.5 $\pm$ 0.1 (15)	11.6 $\pm$ 0.1 (18)
Control	13.9 $\pm$ 0.1 (81)	13.9 $\pm$ 0.1 (81)	13.9 $\pm$ 0.1 (81)

Values shown are the mean day of eyelid opening $\pm$ s.e.; values in parentheses are the numbers of mice treated at each dose level. The doses are per g body weight. The data from four experiments, all of which yielded similar results, were pooled. Newborn mice were treated with growth factors dissolved in PBS or with PBS alone, as described in the text. The doses for the three growth factors were adjusted to be equivalent to a murine EGF standard, as measured¹⁸ by the ability of each peptide to compete with ¹²⁵I-labelled EGF for binding to cell-surface receptors in normal rat kidney fibroblasts (NRK 49F). Based on amino-acid analysis, the amount of TGF- α per binding equivalent was fourfold greater than that of either mouse or human EGF. The effects of all dosages of TGF- α and both EGFs are significantly different from the controls (Student's *t*-test; *P* < 0.001). There are no significant differences between the effects of the three peptides themselves at any given dose; *P* > 0.5.

was purified by passage through Sephadex G-75 and HPLC after expression of the cloned gene in *E. coli*⁶; the material used was essentially free of other proteins, as shown by the absence of isoleucine and methionine on amino-acid analysis. Human EGF was obtained by chemical synthesis of the gene for this peptide¹⁷, followed by expression of the gene in *E. coli*¹⁷. The expressed precursor protein (provided by G. D. Searle) was cleaved with trypsin to give the exact human sequence, and the cleaved protein was isolated by ion-exchange and gel chromatography. The purity of this peptide was demonstrated by HPLC in three systems, by acrylamide gel electrophoresis, and by the absence of threonine and phenylalanine on amino-acid analysis. Mouse EGF was prepared by acid/ethanol extraction of salivary glands, followed by gel filtration, DEAE-cellulose chromatography, and HPLC¹⁸, which is a modification of the procedure of Savage and Cohen¹⁹. The purity of the murine EGF was shown by the absence of alanine, lysine and phenylalanine from the final material.

Newborn Swiss mice were injected daily, beginning at 1 day old, with one of the three growth factors or with a PBS control. All injections were made subcutaneously, in the nape of the neck, in a volume of 10 µl per g body weight. Each litter of nine mice was divided into three groups, which received different doses of the same growth factor, or the same dose of each of the three growth factors. Each mouse was treated for eight consecutive days; the doses used are shown in Table 1. Figure 1 and Table 1 show that human TGF- α , and human and murine EGF all accelerate normal eyelid opening, and that the effect is dose-responsive, except at the very lowest dose used. Moreover, the response of the newborn mouse to the three different peptides is essentially identical.

In another set of experiments (data not shown) we found that transforming growth factor- β (TGF- β), isolated from human blood platelets²⁰ and used at a dose of 400 ng per g body weight, either alone or in combination with murine EGF, had no activity in the eyelid opening assay. TGF- β has been shown to be a potent synergist of EGF in promotion of the growth of non-neoplastic rat fibroblasts in soft agar^{7,12,20}.

Although the activity of human EGF in this eyelid opening assay has been described previously in tests involving small numbers of mice^{2,3}, the present data show definitively for the first time that the respective human and murine peptides are essentially identical in this regard. Furthermore, the almost identical activity of TGF- α and the two types of EGF suggests that the *in vivo* response to all three peptides involves a common, receptor-mediated pathway, much as has been shown for their effects *in vitro*. The equivalence of these peptides *in vivo* is all the more remarkable considering that pharmacokinetic phenomena such as absorption, transport and half-life are also involved. Finally, the data are of interest in that they show that a defined peptide growth factor, generally associated with the malignant behaviour of cells, may have a benign physiological activity that is indistinguishable from related molecules found in non-malignant tissues.

We thank Stanley Cohen for originally suggesting these experiments, Richard Assoian for amino-acid analysis of the three peptides, and Stephen Dalton and Sue Perdue for secretarial assistance.

Received 19 February; accepted 2 April 1985.

1. Cohen, S. *J. biol. Chem.* **237**, 1555-1562 (1962).
2. Gregory, H. *Nature* **257**, 325-327 (1975).
3. Cohen, S. & Carpenter, G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1317-1321 (1975).
4. Marquardt, H., Hunkapiller, M. W., Hood, L. E. & Todaro, G. J. *Science* **223**, 1079-1082 (1984).
5. Marquardt, H. & Todaro, G. J. *J. biol. Chem.* **257**, 5220-5225 (1982).
6. Derynck, R., Roberts, A. B., Winkler, M. E., Chen, E. Y. & Goeddel, D. V. *Cell* **38**, 287-297 (1984).
7. Roberts, A. B., Frolak, C. A., Anzano, M. A. & Sporn, M. B. *Fedn Proc.* **42**, 2621-2626 (1983).
8. De Larco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4001-4005 (1978).
9. Todaro, G. J., Fryling, C. & De Larco, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5258-5262 (1980).
10. Sporn, M. B. & Todaro, G. J. *New Engl. J. Med.* **303**, 878-880 (1980).
11. Sporn, M. B. & Roberts, A. B. *Nature* **313**, 745-747 (1985).
12. Anzano, M. A. *et al. Cancer Res.* **42**, 4776-4778 (1982).
13. Massagué, J. *J. biol. Chem.* **258**, 13606-13613, 13614-13620 (1983).

14. Twardzik, D. R., Todaro, G. J. & Marquardt, H. *Science* **216**, 894-897 (1982).
15. Cohen, S. & Carpenter, G. A. *Rev. Biochem.* **48**, 193-216 (1979).
16. Hollenberg, M. D. *Vitam. Horm.* **37**, 69-110 (1979).
17. Smith, J. *et al. Nucleic Acids Res.* **10**, 4467-4482 (1982).
18. Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M. & Sporn, M. B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5339-5343 (1981).
19. Savage, C. R. & Cohen, S. *J. biol. Chem.* **247**, 7609-7611 (1972).
20. Assoian, R. K., Komoriya, A., Meyers, C. A., Miller, D. M. & Sporn, M. B. *J. biol. Chem.* **258**, 7155-7160 (1983).

Expression of *c-mos* proto-oncogene transcripts in mouse tissues

Friedrich Propst & George F. Vande Woude

Molecular Mechanisms of Carcinogenesis Laboratory, LBI-Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

Valuable information about proto-oncogenes and their physiological function has been obtained by studying their expression in normal cells. However, expression of the *c-mos* gene, the cellular homologue of the transforming gene of Moloney murine sarcoma virus^{1,2}, has not been detected in normal mouse cells or tissues³⁻⁶. The conservation of the *c-mos* open reading frame⁷⁻⁹ strongly indicates that the gene must function during some portion of the animal life cycle, and other lines of evidence¹⁰⁻¹⁵ suggested to us that the *c-mos* proto-oncogene may be expressed at very low levels in normal tissues. We have used a sensitive *S*₁ nuclease assay to screen RNA preparations from mouse tissues and describe here the detection of *c-mos*-related transcripts especially in mouse embryos, testes and ovaries. The transcripts found in testis RNA are estimated to be ~1.7 kilobases (kb) long by Northern analysis. *S*₁ analysis demonstrates that the entire *mos* open reading frame is present. In contrast, we detect ~1.4-kb transcripts in ovary RNA and at least two major transcripts, ~2.3 and ~1.3 kb, in embryo RNA. The latter transcripts have in common sequences of at least 1 kb, representing most of the *c-mos* open reading frame. The variation in size of the *mos* transcript in different tissues suggests a novel regulatory mechanism for the expression of this proto-oncogene.

In the present study, we have examined mouse tissues for *c-mos* expression using *S*₁ nuclease^{16,17}. An internal fragment

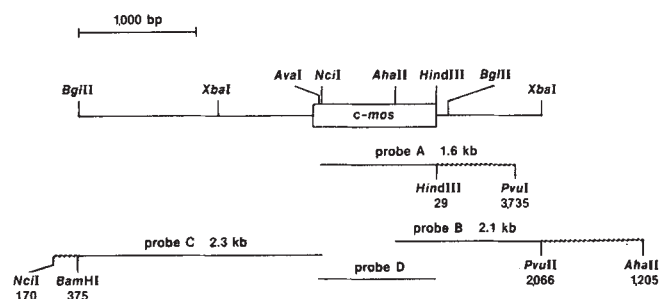
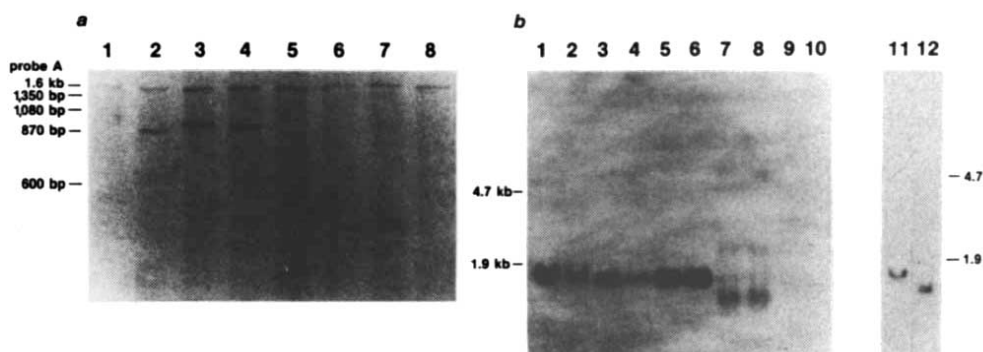


Fig. 1 Restriction fragment probes derived from the mouse genomic *c-mos* locus. Relevant restriction sites are shown. Wavy lines represent pBR322 sequences (restriction sites and their sequence position are given below the line). Probe A, 3' end-labelled at *Ava*I; probe B, 3' end-labelled at *Aha*II; probe C, 5' end-labelled at *Nci*I; and probe D, *Ava*I/*Hind*III fragment labelled by nick-translation.

Methods. All probes were derived from appropriate subclones of *c-mos* sequences in pBR322; 3' end-labelling was carried out with the Klenow fragment of *Escherichia coli* DNA polymerase (Boehringer) in the presence of the appropriate [α -³²P]-labelled deoxynucleotides (3,200 Ci mmol⁻¹; NEN) required to fill-in the respective restriction site as described elsewhere²⁰, except that labelling was done in the presence of 240 mM NaCl. 5' End-labelling was carried out using polynucleotide kinase (Boehringer)^{20,23} and [γ -³²P]ATP (7,000 Ci mmol⁻¹, NEN). The *Ava*I/*Hind*III fragment for probe D was gel-purified and nick-translated as described elsewhere²⁴.

Fig. 2 Detection of transcripts homologous to *c-mos*. **a**, S_1 analysis using probe A. Lane 1, 25 μ g of yeast transfer RNA; lane 2, 2.5 μ g of total RNA from a NIH 3T3 cell line transformed by pm13 plus 25 μ g of total NIH 3T3 cell RNA; lane 3, 50 μ g of total embryonic RNA; lane 4, 5 μ g of embryonic poly(A)⁺ RNA plus 25 μ g of yeast RNA; lanes 5–8, 25 μ g of total RNA from, respectively, liver, kidney, brain and skin. The samples were hybridized to probe A at 60 °C overnight as described^{16,17}, followed by digestion at 37 °C for 45 min^{16,17} with single-strand-



specific S_1 nuclease (10,000 U ml⁻¹ final concentration; Boehringer). Samples were then analysed by electrophoresis on a 5% denaturing polyacrylamide gel, which was then dried and exposed to X-ray film (XAR-5; Kodak) at -70 °C for 20 h in the presence of an intensifying screen. In this analysis a small amount of self-annealing probe is detected at 1.6 kb. **b**, Northern analysis using probe D. Lane 1, 25 μ g total testis RNA; lane 2, 25 μ g testis RNA (fraction not bound to oligo(dT) cycle 1); lane 3, 2 μ g testis RNA (fraction bound, cycle 1); lane 4, 2 μ g testis RNA (fraction not bound, cycle 2); lane 5, 2 μ g testis RNA (fraction bound, cycle 2); lane 6, 2 μ g testis RNA (fraction bound, cycle 3); lane 7, 10 μ g male embryo poly(A)⁺ RNA; lane 8, 10 μ g female embryo poly(A)⁺ RNA; lane 9, 25 μ g total male embryo RNA; lane 10, 25 μ g total female embryo RNA; lane 11, 25 μ g total testis RNA; lane 12, 25 μ g total ovary RNA. Samples 1–10 and 11–12, respectively, were electrophoresed on separate agarose formaldehyde gels and transferred to nitrocellulose as described elsewhere^{25–27}. Hybridization was carried out at 42 °C in 50% formamide²⁰, then the blots were exposed to X-ray film (XAR-5; Kodak) for 3 days. RNA preparation: Animals were killed by cervical dislocation and the various organs were removed. Embryos for the experiment in **a** (and in Fig. 3a) were a mixture from nine different inbred strains removed 1 day before birth. The BALB/c embryos for the experiment in **b** were removed at day 19 of gestation. In each case the embryo was completely separated from extra-embryonal membranes and the placenta. For the experiment in **b**, embryos were sexed by microscopic examination of the gonads. Testes and ovaries were taken from 34–39-week-old BALB/c breeders. Routinely, specific organs from several animals were pooled before homogenization, for example, testes from 10–15 mice. The tissue was homogenized immediately or frozen on dry ice and kept at -70 °C for later homogenization (using a mortar and pestle) to a fine powder. Fresh tissue and the frozen powder were then homogenized in a Dounce homogenizer in 5 M guanidinium isothiocyanide, 0.1 M mercaptoethanol and total RNA was prepared as described elsewhere^{20,28,29}. In **a**, poly(A)⁺ RNA was prepared as described elsewhere²⁰; the same poly(A)⁺ RNA selection procedure was used for **b**, except that oligo(dT) columns were not washed with low-salt buffer after loading.

of mouse *c-mos* was radioactively end-labelled with all four deoxynucleotides and used as a probe (probe A, Fig. 1). This probe was then used to screen for *mos*-containing transcripts in total or poly(A)⁺ RNA extracted from a variety of adult mouse tissues and embryos (Fig. 2a). As a positive control for the S_1 assay, we used total RNA extracted from NIH 3T3 cells transformed by pm13, a recombinant DNA clone that contains *v-mos*; this cell line was estimated to contain 1–10 copies of *v-mos* RNA per cell^{10,11}. *mos*-Specific transcripts were detected in a 10-fold dilution of total RNA prepared from these transformed cells (Fig. 2a, lane 2).

Probe protection was clearly observed in RNA prepared from mouse testes, ovaries (see below) and embryos (Fig. 2a, lanes 3 and 4), the size of the protected fragment corresponding to the distance between the *Ava*I and *Hind*III sites (925 base pairs, bp) in *c-mos*. RNA from pm13-transformed NIH 3T3 cells protects a somewhat smaller fragment (Fig. 2a, lane 2) because the 3' end of this particular *v-mos* gene differs from that of *c-mos*^{18,19}. To exclude the possibility that contamination by DNA of total RNA, especially testis RNA, was responsible for protection of the 925-bp fragment, we performed reconstruction experiments with unlabelled homologous *c-mos* DNA fragments (not shown) or subjected samples to digestion with DNase or RNase before the assay. The hybridization conditions were chosen to favour RNA/DNA hybrid formation (as opposed to DNA/DNA hybrids). RNase but not DNase treatment prevented the detection of the ~925-bp band (not shown). We have variably observed protection of the ~925-bp *c-mos* band in S_1 analyses of total RNA from epididymis, kidney, brain and placenta, but the low amounts detected in these preparations were not analysed further. We did not detect similar bands in total RNA prepared from liver or skin (Fig. 2a, lanes 5 and 8), spleen, thymus and intestine (not shown).

We observed that the concentration of *c-mos* RNA in testes, ovaries and embryos was equal to or higher than that in the pm13 control (compare lanes 1, 2 and 3 of Fig. 2a), suggesting that we should be able to identify the transcripts by Northern analysis. We detected discrete bands of ~1.7 kb in testis RNA

(Fig. 2b, lanes 1–6 and 11), 1.4 kb in ovary RNA (Fig. 2b, lane 12) and ~1.3 and ~2.3 kb in embryonic poly(A)⁺ RNA (Fig. 2b, lanes 7, 8). In these analyses, the poly(A)⁺ RNA was prepared from either male (Fig. 2b, lane 7) or female (Fig. 2b, lane 8) embryos.

Comparison of the concentration of *mos* transcripts in total embryo RNA versus poly(A)⁺ RNA (Fig. 2a, lanes 3 and 4) indicates that the oligo(dT) selection does not result in a significant enrichment of *mos* RNA. However, the same selection procedure²⁰ does enrich for *c-myc* transcripts (not shown). Even with decreased stringency conditions for poly(A)⁺ RNA selection, a significant amount of *mos*-specific RNA is not bound to oligo(dT) (Fig. 2b, lanes 1 and 2). However, the fraction of *mos* transcripts which are bound in the first oligo(dT) selection (Fig. 2b, lane 3) continues to be selected through two additional cycles (lanes 4–6 of Fig. 2b); this suggests that the *mos* transcripts may have only short poly(A) tails and may help to explain some of the previous difficulties in detecting *mos* transcripts.

To characterize further the *mos* RNA expressed in testis and embryonic tissue, probes B and C (Fig. 1) were used with S_1 nuclease to determine the extent of protection of sequences downstream and upstream of the *c-mos* open reading frame. With both testis and embryonic RNA, a ~470-bp fragment from probe B was protected (Fig. 3a, lanes 1, 2 and 3 respectively). This result indicates that *c-mos* RNA extends 130 bp downstream from the *c-mos* termination codon and ends ~20 bp beyond a consensus poly(A) signal⁹. S_1 mapping of testis RNA with probe C yields two bands of ~400 and 450 bp (Fig. 3b, lane 1) corresponding to sites 260 and 310 bp upstream of the conserved *mos* initiation codon²¹. The DNA sequence in this region shows no obvious splice acceptor or promoter sites. However, it is just distal to a region called MUH²², which is 75% conserved between human and mouse. These analyses demonstrate that the entire *mos* open reading frame plus 260 or 310 bp of upstream sequence and 130 bp of downstream sequence are present in the transcripts. The estimated length of the testis transcript is 1.7 kb by Northern analysis, but only ~1.4 kb can be accounted for by S_1 mapping. Thus, either the

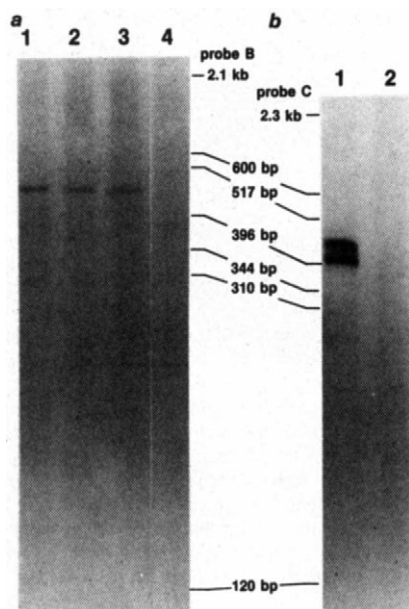


Fig. 3 S₁ nuclease mapping of *mos* transcripts. **a**, S₁ mapping using probe B. Lane 1, 25 µg total testis; lane 2, 2 µg of poly(A)⁺ testis RNA plus 30 µg of yeast tRNA; lane 3, 3 µg of embryonic poly(A)⁺ RNA plus 30 µg of yeast tRNA; lane 4, 30 µg of yeast tRNA. The samples were hybridized to probe B at 60 °C overnight. **b**, S₁ mapping using probe C. Lane 1, 25 µg of total testis RNA; lane 2, 30 µg of yeast tRNA. The samples were hybridized to probe C at 52 °C overnight. For analysis of DNA/RNA hybrids and preparation of RNA samples, see Fig. 2 legend. The X-ray film (XAR-5; Kodak) was exposed for 14 days at -70 °C in the presence of an intensifying screen.

transcript size is overestimated by Northern analysis or there is RNA processing 5' and/or 3' to the protected regions.

While the embryonic RNA transcripts protect a discrete ~925-bp fragment of probe A (Fig. 2a, lanes 3 and 4) and a ~480-bp fragment of probe B (Fig. 3a, lane 3), they do not give a detectable signal with probe C (not shown), indicating that they differ from the testis RNA transcripts in the region 5' to the *mos* ATG. Similarly, S₁ mapping of the ovarian RNA transcripts indicates that they also differ from both the testis and embryo transcripts at the 5' end (not shown). Collectively, these results suggest that transcripts from the single-copy *mos* locus are expressed in a tissue-specific manner. In addition, these transcripts are either promoted from different sites and/or are spliced differently in testes, ovaries and embryonic tissues.

We have shown that *c-mos* is not efficiently activated by a downstream long terminal repeat (in contrast to *v-mos*¹); this has been attributed to the presence of an upstream transcription termination sequence termed UMS²¹ (M. L. McGeady and G.F.V.W., in preparation) and to the lack of a promoter between UMS and *c-mos* that can function in NIH 3T3 cells²¹. How then can *mos* be expressed? One suggestion²¹ is that *mos* expression can be *trans*-activated. The presence of *mos* transcripts in tissues like testes and ovaries may indicate a *trans*-activation mechanism mediated by hormonal control. We also note that an additional 60 codons are proximal to the conserved *mos* ATG in all three *mos* genes that have been sequenced^{8,22}. Thus, in addition to tissue-specific regulation of the size of the *mos* transcripts expressed, the presence or absence of these codons (or possibly other exons) could give rise to functionally different protein products in specific tissues.

Finally, the higher levels of *mos* expression detected in gonadal tissues and the differences in transcript size between ovaries and testes suggest that *mos* proto-oncogene expression is regulated in a sex-linked manner and may be associated with expression of primary and/or secondary sex-related phenotypes.

This research was sponsored by the National Cancer Institute, DHHS, under contract NO1-CO-23909 with Litton Bionetics, Inc.

Note added in proof: We have also detected a 6-kb *mos* transcript in poly(A)⁺ RNA from embryos and epididymis. In addition, *mos* transcripts are present in rat embryos and testes.

Received 4 March; accepted 9 April 1985.

- Oskarsson, M., McClements, W. L., Blair, D. G., Maizel, J. V. & Vande Woude, G. F. *Science* **207**, 1222-1224 (1980).
- Jones, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2651-2655 (1980).
- Gattoni, S., Kirschmeier, P., Weinstein, I. B., Escobedo, J. & Dina, D. *Molec. cell. Biol.* **2**, 42-51 (1982).
- Muller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. M. *Nature* **299**, 640-644 (1982).
- Goyette, M., Petropoulos, C. J., Shank, P. R. & Fausto, N. *Molec. cell. Biol.* **4**, 1493-1498 (1984).
- Maguire, R. T., Robins, T. S., Thorgerisson, S. S. & Heilman, C. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1947-1950 (1983).
- Watson, R., Oskarsson, M. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4078-4082 (1982).
- van der Hoorn, F. A. & Firzlauff, J. *Nucleic Acids Res.* **12**, 2147-2156 (1984).
- Van Beveren, C., van Straaten, F., Galleshaw, J. A. & Verma, I. M. *Cell* **27**, 97-108 (1981).
- Wood, T. G., McGeady, M. L., Blair, D. G. & Vande Woude, G. F. *J. Virol.* **46**, 726-736 (1983).
- Wood, T. G., Blair, D. G. & Vande Woude, G. F. in *Perspectives on Genes and the Molecular Biology of Cancer* (eds Roberson, D. L. & Saunders, G. F.) 299-306 (Raven, New York, 1983).
- Rechavi, G., Givol, D. & Canaani, E. *Nature* **300**, 607-611 (1982).
- Gattoni-Celli, S., Hsiao, W.-L. W. & Weinstein, I. B. *Nature* **306**, 795-796 (1983).
- Cohen, J. B., Unger, T., Rechavi, G., Canaani, E. & Givol, D. *Nature* **306**, 797-799 (1983).
- Papko, J., Verma, I. M. & Hunter, T. *Cell* **29**, 417-426 (1982).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
- Favaloro, J., Treisman, R. & Kamen, R. *Meth. Enzym.* **65**, 718-749 (1980).
- Donoghue, D. J. & Hunter, T. *J. Virol.* **45**, 607-617 (1983).
- Brow, M. A. D., Sen, A. & Sutcliffe, J. G. *J. Virol.* **49**, 579-582 (1984).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (eds). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Wood, T. G., McGeady, M. L., Baroudy, B. M., Blair, D. G. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7817-7821 (1984).
- Blair, D. G. *et al. in Cancer Cells: Oncogenes and Viral Genes* Vol. 2 (eds Vande Woude, G. F., Levine, A. J., Topp, W. C. & Watson, J. D.) 281-289 (Cold Spring Harbor Laboratory, New York, 1984).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499 (1980).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237-251 (1977).
- Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. *Biochemistry* **16**, 4743-4751 (1977).
- Rave, N., Crkvenjakov, R. & Boedtker, H. *Nucleic Acids Res.* **6**, 3559-3567 (1979).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Glisin, V., Crkvenjakov, R. & Byus, C. *Biochemistry* **13**, 2633-2637 (1974).
- Ullrich, A. *et al. Science* **196**, 1313-1319 (1977).

Errata

Does interocular transfer occur in visual navigation by ants?

R. Wehner & M. Müller
Nature **315**, 228-229 (1985)

The third sentence of the last column on page 229 should read "In honey bees, IOT has been inferred ..."

Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast

A. M. Miller, V. L. MacKay & K. A. Nasmyth
Nature **314**, 598-603 (1985).

In the eighth line of paragraph two, α/α diploid should read a/α diploid, and in the upper part of Fig. 5a and in the third line of the Discussion, $\alpha 1/\alpha 2$ should read $a1/\alpha 2$.

What's the matter?

Niles Eldredge

The Problems of Evolution. By Mark Ridley.

Oxford University Press: 1985. Pp. 159. Hbk £12.50, \$19.95; pbk £3.95, \$8.95.

PROBLEMS in biology, in sharp distinction to those of philosophy, are solvable — or so writes Mark Ridley in the preface to *The Problems of Evolution*. Ridley has organized his book around a series of ten Great Problems currently (albeit, he says, timelessly) besetting evolutionary biology, problems that curiously decrease in their apparent solvability quotient as one ascends the scale of biological complexity and at the same time approaches the end of the book. Ridley's preoccupation with the answers to questions imparts an eerie quality to his narrative. On the one hand he is often quite sure of having the answer to a particular problem (for example the gene is the unit of selection), a sort of certainty I had thought was no longer fashionable in science in general (and particularly in the domain of evolutionary studies, justly known for its authoritative imaginary tales). On the other hand, I must admit to some frustration when Ridley throws up his hands and announces that a problem still defies solution, and that, in fact, our understanding remains at the same low level attained in earlier days. It seems to me that rather more progress has been made towards resolving some of the confusion in evolutionary biology than Ridley seems willing to concede.

Whence this collision between certitude and confusion? Ridley's book accurately reflects the generally unsettled state of evolutionary biology, for which, I think, David Hull is right to lay the blame on a pervasive "common sense" ontology that lies at the very core of the subject. Thus certitude stems from unexamined convictions on the very nature of such entities as genes, organisms, demes, species and taxa of higher categorical rank; and confusion ensues if we get our ontology wrong. Ridley's discussion of what species are — a stunning rerun of the Dobzhansky-Mayr position enunciated nearly 50 years ago — forms the crucial case in point. In this view, species are seen to be simultaneously reproductive communities and fairly coherent groupings of similar organisms, recognizably distinct from other such groups. Ridley is thus forced to repeat the standard claim that species are "real" at any one time (especially well demarcated in sympatry with close relatives), but through time will, of necessity, disappear, evolving themselves out of existence as they transform into descendants. Species exist only in a single time plane, according to this view, a conceptualization that understandably outraged George Gaylord Simpson.

In this received ontology, species are

real, individual entities when construed as reproductive communities at any one time; they are also classes of similar organisms. Moreover, the economic adaptations of organisms (the main source of that similarity) are imagined to be in a constant state of overhaul, and so the properties of the members of the class change, and a species evolves itself out of existence. Ridley simply does not report that Ghiselin and Hull over a decade ago pointed out that "things" are either classes or individuals, and that there is no real trick to seeing species, if construed as reproductive communities, as individuals in time as well as in space. Species emerge as spatiotemporally bounded historical entities, regardless of how much, or little, adaptive modification may accrue in the phenotypes of their component organisms. Much of the doubt and confusion permeating the second half of this book stems directly from Ridley's certitude over the basic nature of species, based as it is on a rather garbled ontology.

Indeed, overall the book is an excellent statement of mainstream evolutionary biology as it stands today. If it does not expose the *real* problem in evolutionary theory — the ontological issues — it does paint an accurate picture of the teleological-ridden "who or what benefits" approach that selectionists seem ever more locked into as a source of explanation for all manner of biological phenomena. Nor is Ridley utterly conventional. Within the overall matrix of neo-Darwinian theory, he follows the Williams-Dawkins reductionism that sees the gene as the locus of evolutionary action. And, in an amusing and effective gambit that coincides well

with my own prejudices, Ridley discusses systematics by contrasting phenetics and cladistics, claiming the latter is evolutionary and the former not, thereby allowing him to brush aside the vast middle ground of "evolutionary systematics" as an uninteresting muddle of the two extremes. Moreover, while Ridley is clear that cladistics does not depend upon any prior theory of how evolution actually works, in several places he makes a link between nested patterns of resemblance linking up all elements of the biota and the simple notion of descent with modification. Those nested patterns of similarity have always been the strongest evidence that life must have had an evolutionary history, and it is refreshing to see it acknowledged once again, just as a journalist in the United States, in a feat of literary legerdemain, has recently tried to show how cladistics somehow throws doubt on the very idea of evolution.

Ridley's book is not a "fun" read. Indeed, it is difficult to determine to whom it is addressed. The style, superficially zesty ("And the prediction is this . . . Here is the final result"), is actually rather flat and utterly matter-of-fact: there seems for Ridley no joy in these rather marvellous mysteries. The level is decidedly elementary — as if the intended audience were beginning students or the ever-elusive "intelligent layman" — yet the tone is so unrelentingly serious and the pace so brisk that all but the most dedicated will surely flag.

Perhaps most importantly, however, the book does not get at what I think is *really* the matter with evolutionary theory. But as a quick yet pretty accurate summary of most of the current topics under debate, and as a source of insight about how most evolutionists still approach their topic, it should prove useful indeed. □

Niles Eldredge is at the American Museum of Natural History, New York.

New journals review

On 26 September *Nature* will publish the fifth annual review supplement devoted to science journals.

Criteria for inclusion of a journal in the 1985 issue are that:

(i) the first number appeared, or the journal was retitled, between June 1983 and May 1984 (the second cut-off date allows at least three issues of a journal to have been published, the minimum number on which a reasonable judgement can be based);

(ii) it is published at least three times a year;

(iii) the main language used is English.

Publishers and learned societies are invited to send four different issues of each suitable periodical, including the first and most recent numbers (if from outside the United Kingdom, by air mail) to: The Review Editor, *Nature*, 4 Little Essex Street, London WC2R 3LF, England. Subscription details for both 1985 and 1986 should be included.



Home in hyperspace

Deborah Singmaster

The Fourth Dimension and How to Get There. *

By Rudy Rucker.

Rider-Hutchinson: 1985. Pp.228. £9.95.

THE dust jacket of the British edition of Rudy Rucker's *The Fourth Dimension* carries a picture that looks like a shot from the final sequence of *Star Wars*. It could be a post card from the fourth dimension, and if it were from the author it would say "Wish you were here!". Rucker admits that he is in love with cosmology and in this exhilarating book he tries to convey his enthusiasm to his readers.

The book is dedicated to "A Square, on his hundredth anniversary" — an apt tribute to another book on the same theme, *Flatland* by the English school master Edwin A. Abbott, published in 1884 and gallantly kept in print in Britain by Basil Blackwell. It deserves to be more widely read than it is. Rucker makes frequent references to *Flatland* throughout his book; he has even written a sequel to the original.

Starting with Plato and the allegory of the shadows on the wall of the cave, Rucker moves by analogy towards a description of a spatial fourth dimension. He introduces the hypercube, "probably the best-known 4-D geometrical problem", and proceeds to the hypersphere, hyperspace tunnels, the shape of space, spacetime and so on. From these literal interpretations of a fourth dimension, he passes to more abstract uses of the term and suggests that we, with our complex interior world of thoughts, feelings, dreams, are "in some very real sense, beings of more than three dimensions".

Rucker has no difficulty in thinking four dimensionally. He quotes Charles Hinton, "the great hyperspace philosopher" and a contemporary of Abbott's, as saying that one of the best methods of preparing oneself for four-dimensional thought is to create a deliberate confusion between left and right, by experimenting with mirror images, for example. One of the most entertaining of the many anecdotes in the book is Rucker's potted biography of Hinton, the son of James Hinton who advocated and practised free love. Charles, who was himself convicted of bigamy, believed that "space is, at the deepest level, to be known by the heart and not the brain". But that way mysticism lies: Rucker is at times unashamedly mystical.

For mathematical greenhorns, Rucker provides simple explanations of major concepts including Einstein's theory of

relativity, quantum mechanics and David Hilbert's theory of infinite dimensional spaces. He also discusses less orthodox notions such as synchronicity and the possibility of time travel — he is, after all, apart from being a mathematical logician, a writer of science fiction.

Appropriately, *The Fourth Dimension* is a fourfold book. Besides the text there are witty and helpful diagrams and illustrations by David Povilaitis, the margins contain extensive quotations from writers as diverse as Plato, Ouspensky, Lewis Carroll and Borges, and there are also puzzles for those whom Martin Gardner, in his foreword, describes as "the

mathematically competent". (Answers are supplied.)

Rucker's book is not aimed at the trained mathematician or physicist but it should appeal to anyone else who is interested in the nature of the Universe and our part in it. Science undergraduates, even sixth formers, and all science fiction enthusiasts will find it (and *Flatland*) stimulating reading. This is a refreshing and thought-provoking book written by an engaging optimist for whom, unlike Pascal, infinity holds no fears. □

Deborah Singmaster is a freelance writer and broadcaster.

Distant reactions

Alexander Dalgarno

Interstellar Chemistry.

By W.W. Duley and D.A. Williams.

Academic:1984. Pp.251. \$45, £29.

THE discovery in the dense regions of interstellar space of molecules, ranging in complexity from simple diatomic species to extended linear chain polyynes, has provided a unique diagnostic probe into the physical conditions in which star formation occurs. Analysis of the intensities of molecular emission and absorption lines yields estimates of the densities, temperatures, velocities and radiation characteristics of the environment; observations of absorption by rotational levels of the diatomic cyanogen molecule CN have given us a measure of the temperature of the blackbody radiation of the Universe at wavelengths of 2.6mm and 1.3mm; and observations of emission from deuterated molecules have yielded a value for the density of matter in the early Universe, showing that baryonic matter cannot close the Universe.

Molecules are found also in the material ejected by protostellar objects as they begin their evolution to the main sequence, and in the circumstellar shells and stellar winds of massive stars which undergo mass loss throughout their existence. Molecular masers are clear signposts marking phases in the sequence of events that leads to the formation of stars and the dissipation of the remnant material out of which they are made. Vibrationally excited molecular hydrogen is widely observed and provides insight into a diverse range of phenomena, among them galaxies in collision.

Molecules are much more than probes of exotic environments, however; they are also highly effective cooling agents. Molecular emissions carry away thermal energy and bring molecular clouds to the brink of gravitational collapse. Molecules play a controlling role in the ionization balance of collapsing objects and in the dissipation of rotational angular momentum by magnetic braking.

To utilize fully molecular observations in the investigation of astronomical phenomena we must understand the nature of the processes by which molecules are formed and destroyed. Those processes belong to a special kind of chemistry, a chemistry that is now an integral part of astronomy and that is the subject of Duley and Williams's book.

Interstellar chemistry is a combination of gas phase chemistry and surface chemistry. The phenomena under study occur at densities too low for the three-body processes that control terrestrial gas phase chemistry, and the formation of molecular compounds is initiated by radiative associations in which the reactions are stabilized by the emission of photons. The temperatures can be as low as 5K, though in localized regions the passage of a shock heats the gas to several thousand degrees and may dissociate the molecules which then form into new compounds as the heated gas cools and the atoms recombine. The interstellar medium contains grains, refractory particles of dimensions of the order of 10^{-5} cm, on whose surfaces molecules may form to be later released into the gas phase. The gas phase and surface reactions take place together in a medium bombarded by energetic cosmic rays and subjected to X-rays and to ultraviolet photons from embedded stars and exploding supernovae.

Thus a diverse array of basic chemical processes forms the content of interstellar chemistry. Duley and Williams have written a lucid account of these processes in a book in which the background is developed sufficiently to make apparent the relevance of chemistry to astronomy. Here the two disciplines are woven together in a pattern of absorbing interest which should appeal to a broad range of scientists. A limited familiarity with elementary molecular physics is adequate preparation, but at the same time the authoritative tone reflects the fact that both authors are active researchers in this field. Their book will come to be seen as the standard introduction to the subject. □

Alexander Dalgarno is Phillips Professor of Astronomy at Harvard University.

* In the United States *The Fourth Dimension: Toward A Geometry of Higher Reality*, published by Houghton Mifflin. A technical review of the book appeared in *Nature* 312, 214 (1984).

Bone, bones and biomechanics

Charles E. Oxnard

The Mechanical Adaptations of Bones.

By John Currey.

Princeton University Press: 1984. Pp. 294. \$37.50, £37.20.

FOR earlier generations, D'Arcy Thompson's *On Growth and Form* (published in 1917) and P.D.F. Murray's *Bones* (1934) provided the critical descriptions of what was then known about biomechanics and the mechanical adaption of bone. Nearer our own time, F. Gaynor Evans's *Stress and Strain in Bone* (1957) both provided a summary of the exciting if somewhat confused earlier investigations and marked the beginning of new research based upon the insights of modern engineering. John Currey's book is the natural current successor, together with McNeill Alexander's 1968 and 1984 editions of *Animal Mechanics*, and Wainwright, Biggs, Currey and Gosline's *Mechanical Design in Organisms* (1976). Each of these books supplies its own commentary on the general subject, but, although there is inevitably some overlap, they concentrate upon different facets of the field and are aimed at different readerships.

Much of the early work, often dating from the last century and important though it is, confused stress, strain and architecture; the new investigations described by Currey do not. The old studies followed engineering strength-of-materials traditions by presenting biological materials, especially bone, as though they were homogeneous, isotropic and perfectly elastic bodies capable of being described through infinite beam theory. Currey discusses the new approaches which acknowledge that, like many non-organic materials, bone is inhomogeneous, non-isotropic and demonstrates imperfectly elastic, poroelastic and plastic-elastic properties, and each at several different microscopic levels. In particular, these approaches imply that stress in bones, especially irregular bones and the odd-shaped ends of long bones, is better characterized through techniques such as finite element analysis.

Currey recognizes that all bone and all bones are not alike, and gives many examples of how different types of bone and different designs of bones serve differing functions. He discusses, for instance, differences in form and physical properties of bone ranging from the brittleness and stiffness of the auditory elements in whales related to the transmission of sound in water to the curvatures of finger bones used to suspend the bodies of brachiating apes from branches.

In contrast to many other writings on this subject, Currey also describes not only

immediate and ontogenetic bony adaptation, but takes into account the evolutionary perspective. He makes a convincing case for the design of bone by natural selection while recognizing phylogenetic constraints; and he considers the all-pervasive action of natural selection in producing optimum structure while emphasizing that all mechanical properties of biological materials are a compromise between conflicting requirements.

Perhaps most importantly of all, however, in this book Currey provides an introduction for biologists to the basic ideas of strength of materials and mechanics. This is a summary that should do much to remove misconceptions that still remain in the minds of some students and even some investigators in this area. Among them are the idea that if one set of stress trajectories is compressive then the other must be tensile (which does not necessarily follow), that a limb under tension implies bone under tension (which it

need not) and that mechanical stresses are chiefly borne by bone (they are not: the little known interior complexity of the muscle-tendon unit, and those even less understood and poorly described elements, fascial sheets and ligaments, are all important stress-bearing features especially in the larger land animals).

Finally, although Currey's book unhesitatingly simplifies the ideas of strength of materials in order to aid student understanding, it also reminds us, at almost every step, that real biological situations are much more complex; therein lies the future of these studies. The book is, thus, an excellent source of facts and ideas for students and a resource for their teachers, while at the same time providing a superb set of problems for investigators. □

Charles E. Oxnard is University Professor and Professor of Anatomy and Biology at the University of Southern California. He is author of *The Order of Man* (Yale University Press, 1984).

Light on polymers

Herbert Morawetz

Polymer Photophysics and Photochemistry: An Introduction to the Study of Photoprocesses in Macromolecules.

By James Guillet.

Cambridge University Press: 1985. Pp. 391. £45, \$79.50.

IN these days of extreme specialization, it is rare to find an author who is equally at home in photophysics and photochemistry. Thirty years ago James Guillet worked under Norrish's direction on a doctoral thesis dealing with the photodegradation of polymeric ketones. Since then, he has not only continued to be a leader in polymer photochemistry, but has made seminal contributions to the study of fluorescence and phosphorescence of polymers, both in solution and in bulk, and has published stimulating accounts of the "antenna effect" in which the energy absorbed by a small number of chromophores is relayed along a polymer chain to excite a group far removed from the absorption site. While Professor Guillet's hope that such a process would lead to a chemical reaction has not been realized so far, it is an exciting possibility which may well be demonstrated in the future.

Although areas in which Professor Guillet has been active are stressed more heavily in this his latest book, it does contain an excellent coverage of the field. The only major area I missed is the extensive Russian contribution to the study of polymers by emission anisotropy, which has been summarized by Anufrieva and Gotlib (*Adv. Polym. Sci.* 40, 1; 1981). The effect of the adsorption of fluorescent dyes on their emission characteristics might also have been mentioned.

While the book is an outstanding introduction to photophysics and photochemistry for polymer scientists, I am less sure that the discussion of the principles of polymer science will be adequate for those unfamiliar with this field. There are a number of errors: the expansion coefficient α (p.33) is not treated as chain length dependent; rotational isomeric state calculations (p.34) depend little on the height of rotational barriers; there are no D and L configurations in stereoregular vinyl polymers; the lamellar thickness of folded polymer chains (p.46) is of the order of 10 nm not 1000 nm; it is not possible to estimate the intrinsic viscosity from measurements at a single concentration without knowledge of the "Huggins constant" (p.74); the spacing of phenyl groups in the helix of isotactic polystyrene is not "of the order of 3 Å" but much too large for excimer formation (p.178); and Haas *et al.* and Sisido (p.230) did not study cyclization. It is also a pity that the references only run up to 1981. But these are minor blemishes on a most useful book which is a valuable addition to the polymer literature. □

Herbert Morawetz is a Professor in the Department of Chemistry at the Polytechnic Institute of New York.

New in paperback

● A single-volume edition of Christian de Duve's *A Guided Tour of the Living Cell*. Publisher is W. H. Freeman, price is \$33.95, £24.95. Sterling price of the hardback is £35.90 per two-volume set, not £39.90 as printed in the review (*Nature* 314, 694; 1985).

● *Radiative Processes in Astrophysics* by George B. Rybicki and Alan P. Lightman. Publisher is Wiley, price is \$26.50, £23. For review see *Nature* 289, 729 (1981).

● *In Search of Mind* by Jerome Bruner. Publisher is Harper & Row, price is \$6.95, £7.75. For review see *Nature* 308, 779 (1984).

Scattering sources

R.A. Cowley

Theory of Neutron Scattering from Condensed Matter, Vols 1 and 2.

By Stephen W. Lovesey.

Clarendon: 1984. Vol. 1 pp.329. Vol. 2 pp.344. £38, \$59 each volume.

NEUTRON scattering now plays an essential role in our understanding of many properties of condensed matter, and is used in the study of atomic structures, interatomic forces, magnetism, liquids and polymers. Stephen Lovesey has made significant contributions on the theoretical side of the subject. With Walter Marshall, in 1971 he published *Theory of Thermal Neutron Scattering*, which has been an essential reference book for neutron scatterers, particularly in the field of magnetism, and which was influential in the development of the use of correlation functions for describing the scattering.

The present two volumes extend the coverage of the field beyond that of the earlier book; Vol. 1 describes elastic nuclear scattering, scattering by lattice vibrations and by liquids, and some chemical applications, while Vol. 2 deals with elastic and inelastic magnetic scattering. Although the second volume is entirely rewritten, most of its contents and the manner of treatment will be familiar to

readers of *Theory of Thermal Neutron Scattering*.

In neutron scattering, the Van Hove correlation function is a natural bridge between experiment and theory. The task of the experimentalist is frequently to perform measurements of the function which can then be interpreted in terms of a model for the particular system. Unfortunately there is very little in either of the books about the theory of how to obtain it from measurements of the scattered intensity. There is also nothing on the theory of the experimental resolution or other experimental constraints. The books are a helpful source for the calculation of the Van Hove function for a wide variety of soluble or nearly soluble models, and although I found them difficult to read

they will undoubtedly prove valuable for detailed derivations of particular models. It is, however, a pity that the importance of the formulae produced are not illustrated through more examples of experimental results and discussion of when the simple models apply or do not apply to real systems.

Stephen Lovesey has again provided us with a very useful work of reference. These two volumes should not be used as textbooks because of the neglect of experimental aspects of the subject and of applications to real systems. The books, however, will be consulted by neutron scatterers for many years. □

R.A. Cowley is Professor of Physics at the University of Edinburgh.

Masters of swing

Russell H. Tuttle

The Lesser Apes: Evolutionary and Behavioural Biology.

Edited by Holger Preuschoft *et al.*

Edinburgh University Press/Columbia University Press: 1984. Pp.709. £65, \$56.

DESPITE their modest size and colloquial label, the lesser apes (Hylobatidae) can claim a fair share of superlatives. They are the most numerous and variegated, relatively longest limbed, "swingingest", most naturally bipedal, most musical and most socially cohesive of the apes. We do not know whence they came or how they acquired their remarkable adaptive complexes; but few scientists would deny that we are blessed to have them grace the Earth.

Preuschoft and his colleagues have paid them just homage in this well-produced, fact-packed volume, which presents the augmented results of a conference held near Ulm, West Germany, in 1980. It will be a long-lasting source book on *Hylobates*. The organization of the work reflects its major purpose, that is to impress the world with the need to conserve forest habitats and thereby preserve the glorious gibbons of south-eastern Asia. Hence the first of the five sections is devoted to conservation. Its authors are soberly persuasive.

The plights of Javan silvery gibbons, Mentawai Kloss gibbons, South Asian hoolocks, Thai pileated gibbons and Chinese concolor gibbons are particularly acute. Our knowledge about the effects of high-technology warfare on wildlife in Kampuchea, Laos and Vietnam ranges from scanty to nil, but if the destruction of human life is any index the hylobatids probably have only a poor chance of long-term survival there.

It is now virtually certain that monogamy and territoriality are universal among the Hylobatidae, though we cannot yet tell

whether new groups are formed by the same mechanisms or whether defence of a section of the range serves comparable adaptive functions in all species. All of them call dramatically (and beautifully) but there is notable interspecific variation in which of the sexes calls, displays athletically, and chases and attacks in reaction to the calls of neighbours, actual intrusions or other factors. The interplay between calling, monogamy, subsistence and territorial defence in *Hylobates* has proved an attractive problem for a host of biologists, and Sections 3 and 4 of *The Lesser Apes* contain a wealth of information on the topic. Given their relevance to salient problems in theoretical biology, anthropologists can no longer be viewed as the proprietors of the hylobatid apes.

Functional morphologists have made great strides in elucidating the mechanisms underpinning hylobatid brachiation (arm-swinging), bipedalism and vertical climbing. As with conservation and ecological and behavioural puzzles, the lesser apes have ecumenically challenged Asian, European and North American anatomists and kinesiologists to explain their peculiar positional behaviour. Section 2 of the book is an excellent report of results from modern, technically and mathematically sophisticated, studies on hylobatid subjects and specimens. Here we learn that because of their elongated forelimbs, the movements of slowly brachiating hylobatids correspond reasonably well with the model of a mathematical pendulum, and that the elongation of the forelimbs economizes energy during rapid (ricochetal) arm-swinging.

It is heartening to see a reiteration (on pp. 94-95) of my 1975 model for the stem hominoids as hylobatid-sized, versatile arboreal beasts that were not especially adapted for dramatic brachiation. But I would have been happier had Herr Hollih cited Tuttle among the long list of references that inspired him to adopt it. □

Russell H. Tuttle is Professor of Anthropology at the University of Chicago.

Advertisement

IBM PC & APPLE II USERS

Dramatic advances are being made in scientific tools for personal computers. Jandel Scientific specializes in software and peripherals for:

- Planimetry, digitizing
- Interactive Statistics
- Dynamic System Simulation
- Computer-generated lecture Slides
- And more . . .

Call us and let our technical advisors help you find the product you need. Or write or call for our FREE catalog.

JANDEL SCIENTIFIC

2656 BRIDGEWAY
SAUSALITO, CA 94965

CALL TOLL-FREE 800-874-1888
IN CALIF. CALL 415-331-3022

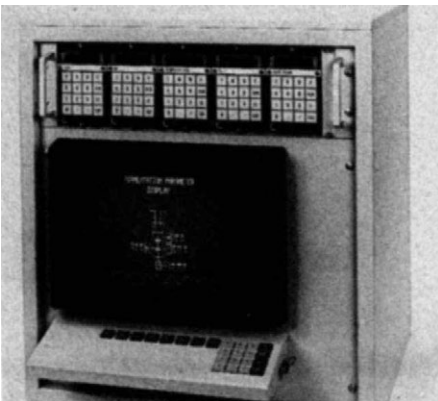
Reader Service No.02

Achema bigger than ever

The Achema-85 exhibition takes place in Frankfurt on 9-15 June. Halls 6-I, -II and -III cover laboratory and analytical techniques, Hall 2 research and innovation.

● The theme of Mettler's exhibit at Achema is laboratory efficiency. Mettler is showing a selection of electronic analytical and precision balances, as well as various systems that can be realized with balances or instruments. Of particular note is a compounding system for chemical or pharmaceutical companies where a number of balances are connected directly to a single computer which controls and monitors the balances. Another system demonstrates the use of lab robots with balances and instruments. Also on view are automatic systems for titration and thermal analysis — the DL40RC MemoTitrator, the DL20 CompactTitrator for routine analyses, and the DL18 Karl-Fischer titrator. The TA3000 thermal analysis system is now equipped with a graphic video screen, and many visitors will also be interested in the RC1 reaction calorimeter which measures the heat flow of chemical reactions and physical transformations.

Stands FG5-9, Hall 6-III & AB2-3, Hall 5 Reader Service No. 100.

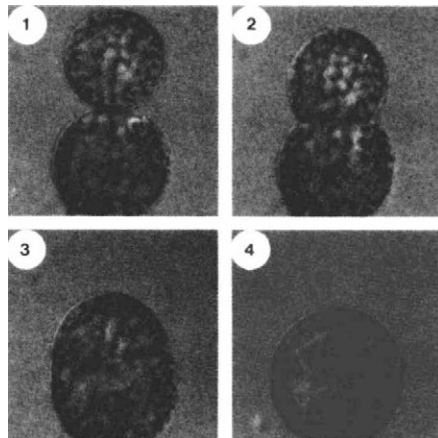


Bioculture control by Applikon.

● The Applikon Bio-Processor ADI 1020 is an advanced and flexible system for total fermentation control. The system is dedicated to control various parameters during the growth of a bioculture, comprising three control levels. Levels 1 and 2 take care of the measurements, controls and data transmission and level 3 is for data handling. The modular build up of levels 1 and 2 allows the system to be tailored to virtually any biological process. At level 3 there is the option either to use the ADI dedicated and preprogrammed unit, including video and keyboard, or the ADI RS 232 interface to connect to a personal computer.

Reader Service No. 101.

These notes are based on information provided by the manufacturers. For further details circle the appropriate numbers on the Reader Service Card bound inside the journal.



Fusion of oat protoplasts by the Krüss electrocell fusion process.

● Krüss GmbH, of Hamburg, makers of equipment to measure surface tension, has moved into genetic engineering with the introduction of the CFA 400 system for electrically induced cell fusion. The latest development from Krüss is an electroporator making it possible to load biological cells with genetic material. In this method, brief electric pulses induce the reversible breakdown of the cell membrane, the cells are "perforated" for a short time without losing viability. Cells (up to 1 ml) are pipetted into a sterile chamber, mixed with the DNA and exposed to a sequence of electric pulses. The DNA is then taken up by the cells. Subsequently the cells are selected for the desired criteria. During electroporation the cells can be observed under a microscope. As well as DNA, other substances, including drugs, may be loaded in cells in this way, with possible implications for specific therapy. The equipment is controlled by a microprocessor and thus easy to operate. Visitors to the Krüss stand will be able to view the fusion of erythrocytes under the microscope on a video system connected to it.

Stand E-19, Hall 6-I

Reader Service No. 102.

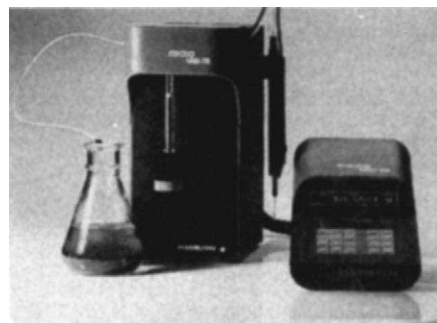
● May & Baker Laboratory Products will be introducing a number of products onto the German market at Achema. On view will be the May & Baker range of HPLC solvents, which includes many solvents that are not available elsewhere. The Pronalys AR range of analytical reagents, Intafaze phase transfer catalysts and Volucon standard concentrates will also be displayed. The Intafaze range includes TDA-1, the first universal crown ether to be made commercially available. TDA-1 is capable of catalysing a wide variety of reactions that would otherwise need expensive solvents.

Stand E22, Hall 6

Reader Service No. 103.

● The Asypor range of membrane liquid filters recently introduced by Domnick Hunter Filters Ltd is claimed to give double the flow rates and double the throughputs achieved with conventional membranes. The Asypor membrane has a unique asymmetrical pore structure. The pore size on the upstream surface is larger than on the downstream surface by a factor of 5 : 1 for optimum performance. Asypor cartridges are available in a range of rated pore sizes from 0.2 to 3 μm with a full range of end cap styles for retrofitting into existing filter housings. The Asypor membrane is also offered in a selection of disks in the standard diameters for process and analytical applications. Asypor disks and cartridges meet FDA requirements and conform to USPXX Class VI. They are integrity testable and have low extractables.

Reader Service No. 104.



Automated dilution and dispensing — the Hamilton Microlab.

● The new Hamilton Microlab M diluter/dispenser should find application wherever precise liquid handling is required for volumes between 1 μl and 25.5 ml. For instance, for sample distribution, dilution series or in the preparation of samples for further processing in different analytical instruments. The Microlab M can also be linked with a balance and a computer for the automatic preparation of standard solutions. Software packages are already available for various balances. The instrument can be controlled by any computer via an RS 232 C interface. Dialogue with the dispensing unit is conducted in ASCII code via the serial data circuit. The Microlab M is programmed via the controller keyboard (up to 587 program steps, maximum storage capacity 99 methods). The stored methods can be viewed ("roll" function) or printed out. Syringe volume, aspirating and dispensing speeds can now be stored, with aspirating and dispensing speeds individually programmable between 1 and 99 seconds.

Stand 19-20, Hall 6-III

Reader Service No. 105.

Advertisements

**SINGER
MICROMANIPULATOR
MK III
Patented**


Perfect control in 3 Dimensions from a single joystick. Continuous variable reductions to suit all magnifications. No after sales maintenance needed.

Details from:
Singer Instrument Co. Ltd.,
Treborough Lodge Roadwater,
Watchet Somerset TA23 0QL
England. Tel. (0984) 40226 Telex:
337492 Comcab G.

Reader Service No.44

LIPOSOMES NEW!
reproducible — homogeneous
— extremely rapid —

The new commercial **LIPOSOMAT®** provides reproducible preparation of uniformly sized unilamellar liposomes (formation within 10 minutes) with very high encapsulation efficiency for both hydrophilic and/or lipophilic agents. Liposome size selectable between 25 and 60nm diameter. Lipid loss far below 1%.

Sample volume normally 5-10 ml or up to 200 ml if used in combination with an additional apparatus.

The **LIPOSOMAT®** is used especially for studying drug/protein/cell — liposome interactions in vitro and in vivo, for reconstitution experiments, and for the transfer of genetic material in gene and biotechnology.

BIANORM-Geräte,
P.O. Box 126, D-8000 MUNICH 65, FRG.

Reader Service No.46

ACADE DIAGNOSTIC SYSTEMS

S.A./N.V., a company specialising in non-isotopic immunoassay using particle counting, is looking for research groups with access to good quality antisera directed towards clinically important assays who would be free to enter into a development contract and/or provide such antisera in commercial quantities.

If interested, please write with full details to: Managing Director, ACADE DIAGNOSTIC SYSTEMS S.A./N.V., Passage de la Vecquée, Box 4100, 1200 Brussels, Belgium. Telephone: 02 762 89 75. Telex: 65867. Replies may be in English, German, French or Dutch.

Reader Service No.72

nature
is available
in microform
from University
Microfilms
International.

Call toll-free 800-521-3044. In Michigan, Alaska and Hawaii call collect 313-761-4700. Or mail inquiry to: University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106.



Schleicher & Schuell's membrane filters and filter holders.

● With the new membrane filters made from Nylon 66, the range of raw materials used in Schleicher & Schuell's membrane filters now includes cellulose nitrate, cellulose acetate, mixed cellulose esters, pure (regenerated) cellulose, as well as polyamide and PTFE. The Nylon 66 membrane filters, suitable for autoclave sterilization, are supplied as circles in diameters from 13 mm up to 293 mm and in pore sizes 0.2 and 0.45 μ m. Nytran membranes are positively charged Nylon 66 membranes for the immobilization of nucleic acids and proteins. They interact with negatively charged biopolymers which become selectively bound and immobilized on the membrane. Because of its high binding capacity and good wet strength, Nytran is an excellent medium for electrophoretic transfer methods in molecular biology. Also on the Schleicher & Schuell stand is an extensive range of products for nucleic acid and protein research, including a number developed during the past two years, such as: DEAE and CM membranes for ionic immobilization, the Minifold II filtration manifold, a further development of the Dot-Blot process (Minifold I), the Quick-Blot kit for the selective immobilization of mRNA and Elutip-d, the mini-column which both purifies and concentrates dsDNA.

Stand DE3-4, Hall 6.

Reader Service No. 106.

● At Achema Elga Ltd, designers and manufacturers of water purification equipment, will be exhibiting a comprehensive range of laboratory and reverse osmosis products. Of particular interest to those involved in HPLC, atomic absorption or tissue culture, will be the updated Elgastat Spectrum RO unit. This system combines the technologies of reverse osmosis, deionization, media and membrane filtration to produce economical reagent grade water which meets all international standards. Elga will also be launching a completely new product for the production of ultra-pure water for use in medical and research laboratories.

Stand E13, Hall 6-II

Reader Service No. 107.

● Boehringer-Mannheim has introduced a new beta-galactosidase substrate, resorufin-beta-D-galactosidase, which has many advantages as substrate in biological systems. The aqueous solution is yellow (λ_{max} 470 nm). The cleavage product after enzymatic hydrolysis is red-violet (λ_{max} 570 nm), and possesses a light-red fluorescence (λ_{max} 583 nm) at pH values higher than 7. The inhibitor programme of Boehringer-Mannheim has been extended with the addition of the protease inhibitors E64 (for thiol proteases), phosphoramidone (for thermolysin), bestatin (for most of the exopeptidases), TLCK (for trypsin) and TPCK (for chymotrypsin), and the new alpha-glucosidase inhibitor bromoconduritol AB. Finally, Boehringer has launched new biochemicals for membrane research. To complement *N*-octylglucoside and *N*-dodecyl-beta-D-maltoside, the new introductions are CHAPS and the peroxide-free tensides Triton X-100, Thesit and isotridecylpolyethylene-glycol ether.

Reader Service No. 108.

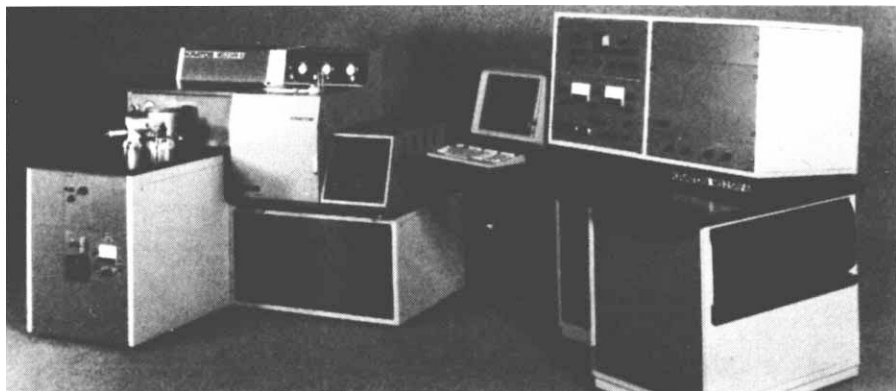


The Bio-Rad Model 620 video densitometer.

● A number of new instruments will be making their debuts on the Bio-Rad stand. Amongst these are a totally digital video densitometer that provides electronic scanning and instant display of data; a portable HPLC analyser designed specifically for on-the-spot applications; a new HPLC MAPS system for the large-scale purification of monoclonal antibodies and a new moderately priced Digilab Fourier transform infrared spectrometer that provides true research-grade performance. Versatile yet simple to use, the Model 620 video densitometer allows the operator to process, optimize, evaluate and view data on a screen before it is printed onto a chart recorder. All operational procedures are selected from a menu displayed on the screen using a mouse-controlled cursor — there is no keyboard. Up to 30 lanes of data can be displayed and printed, and, if required, the operator can arrange these in a number of different formats.

Stand 19-21, Hall 6-III

Reader Service No. 109.



The Kratos MS25 RFA gas chromatography/mass spectrometry system.

● The Kratos Analytical stand features a complete range of liquid chromatography systems, the new MS 25 RFA GC/MS system (also available as an LC/MS), and a series of components for surface analysis. Shown for the first time in Europe is the Spectroflow 950 programmable fluorescence LC detector, which has full microprocessor control of all instrument functions from a touch-sensitive front panel, RS-232 serial interface, or contact closures. DS 650, a chromatography data system with SmartKey control introduced at the Pittsburg Conference, is also on show, together with a range of gradient formers, UV-Vis detectors, post-column reaction systems and pumps. The MS 25 RFA, another instrument on show for the first time in Europe, incorporates inhomogeneous field magnet technology to provide high mass range and ultra-fast scan speeds without loss of spectrometer resolution or sensitivity. An integral DS90 acquisition and processing computer controls the system in conjunction with the spectrometer's "Autoconsole" and distributed microprocessors.

Stand HJ15-16, Hall 6-III

Reader Service No. 110.

● Rheometrics, manufacturer of high technology instruments for material evaluation, has introduced the Rheometrics dynamic analyser (RDA-700). This versatile analytical instrument measures the viscoelastic behaviour of a broad range of materials (including thermoplastics, thermosets, structural composites, rubber, foods, and adhesives) by determining the sample's response to a sinusoidal shearing action. Since the RDA-700 is easy to use and is sensitive to subtle changes in resins and additives, it is ideal for research projects as well as production control. The RDA-700 is a forced frequency instrument rather than a resonant frequency one. This enables the operator to separate the effects of rate and temperature on material properties, making the instrument a sensitive indicator of molecular structure. Rheometrics manufactures a complete line of rheological and impact testing instrumentation for research and development, design, quality control, and process control.

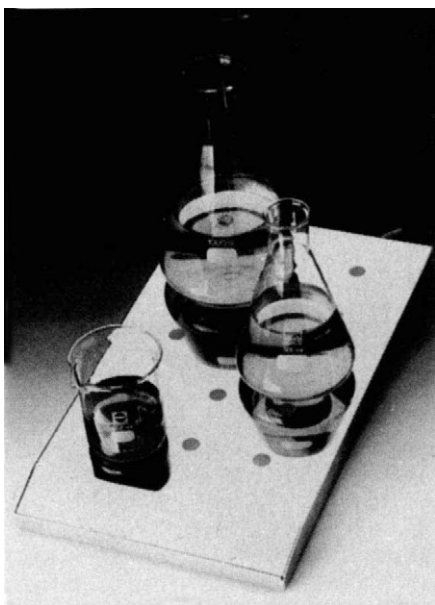
Stand A8, Hall 9-I

Reader Service No. 111.

● The new range of remote-controlled magnetic stirrers from Dr Hoiss + Partner GmbH can be placed under vessels for normal use but can also be used as submersible magnetic stirrers in fluid baths. One possible application would be in sealed or temperature-controlled chambers, where it is advantageous to control the stirring process externally, without opening the chambers. The magnetic stirrers operate most efficiently when they synchronously stir many vessels in the many stirring positions. Different control units are available, suitable for different kinds of stirring tasks; a small, handy plug-in power supply unit or a bench-top unit. Other features include a start-key which simultaneously centres all the stir-bars on the stirring positions, so that it is possible to freely select stir and pause intervals. Following each pause, the stir-bars are re-centred, and the stirring direction reverses. In addition, it is possible to choose the stirring power and thereby facilitate temperature control in closed systems. The compact magnetic drives are available in two models — a 15-position model to simultaneously stir 15 beakers of 250 ml each, or a model for 6 × 1,000 ml Erlenmeyer flasks.

Stand J13, Hall 6-II.

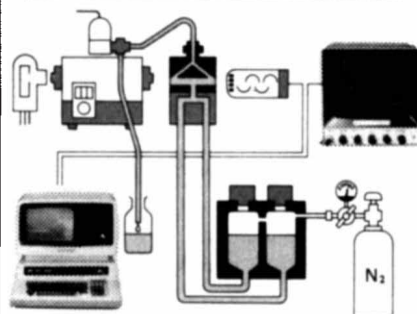
Reader Service No. 112.



Remote-control stirrers from Hoiss + Partner.

Advertisement

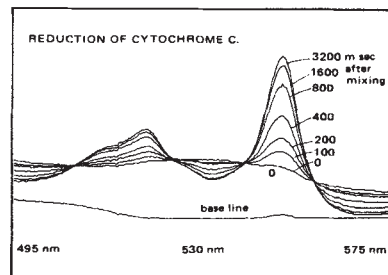
UNION GIKEN First choice for **STOPPED FLOW** reaction analysis



Union Giken apparatus achieves a dead time as short as 1/2 millisecond for measurement of fast reactions. Gas pressure is directly applied to the surface of the reactants in each reservoir, mixing and flow being controlled from the end of the system by a high-response stop valve. The RA-401 STOPPED FLOW SPECTROMETER is the ideal match for fast reactions.



The RA-451 DATA PROCESSOR generates a simulated reaction curve from the measured results and automatically calculates the rate constant.



The RA-415 RAPID SCAN ATTACHMENT monitors the change in absorption spectrum during the reaction.

Other attachments are available for T-jump, P-jump, fluorescence and flash photolysis.

Please visit us Halle 6 III, Stand KL 24 - 25, AICHEMA 85 (ISA).

Contact the sole export agent for more details:—
ATAGO BUSSAN CO., LTD.
7-23, 5-chome Shimbashi, Minato-ku,
Tokyo 105, Japan
Telex: 28421 Cable: TAGOA TOKYO
Phone: (03) 432-8741



Spectrafile — now available for the IBM PC as well as for the Apple IIe.

● Heyden Data systems' stand features a wide range of software for automating laboratory routines using desktop microcomputers, including the IBM PC. Through the cooperation of the instrument manufacturers concerned, Heyden are able to demonstrate the new Sadtler Digital Spectra Libraries on the stand, where Bruker, Digilab and Perkin-Elmer spectroscopy datastations will be operating side by side. The Sadtler collection, formerly in paper volumes, is a comprehensive library of reference spectra. The new digitized collections allow tens of thousands of spectra to be called to the screen. The digitized libraries are also newly available to run on the IBM PC, which can then be used as a stand-alone datastation. Another new development from Heyden display is the IBM PC version of Spectrafile software, available optimized for either IR or UV spectroscopy.

Stand 620/621, Hall 6-III

Reader Service No. 113.



Rapid amino-acid analysis — the Biotronik LC 5001.

● The latest amino-acid analyser from Biotronik, the LC 5001, can analyse physiological fluids in under 2 hours' complete cycle time, and protein hydrolysates in less than 1 hour. Detection levels are 50 picomoles with ninhydrin and 10 picomoles with fluorescence (OPA) detection. Reproducibility is ± 1.5 at 10 nmoles. The instrument uses a microprocessor programming unit which allows 8 different programs to be stored and run in any sequence and combination. Also on the Biotronik stand is the IC 1000 ion chromatograph, now available in two versions. The basic version is for anion determination only and a second version can handle both anions and cations. Full details of Biotronik's user-workshops and applications support services will be available on the stand.

Stand HJ24-25, Hall 6-II

Reader Service No. 114.

● Dyno-Mill, from Willy A. Bachofen AG is the first horizontal, continuous operating agitator ball mill for wet micronization, dispersion and disintegration. In a completely closed system, pumpable suspensions of very low to very high viscosities can be processed. All parts in contact with the product to be ground are easily interchangeable. Models are available with grinding cylinders ranging from 0.3 to 250 litres. A new model in the range, the Dyno-Mill Type KD 20 B, is especially designed for temperature sensitive products and should be suitable for many applications in the pharmaceutical industry and in microbiology. Amongst other items on the stand is the Turbula System Schatz shaker-mixer. This high-performance shaker features three-dimensional overturn motion, with capacities from 2 to 500 litres. *Stand DE-5, Hall 6.*

Reader Service No. 115.

● Amongst the new products to be introduced at Achema by CAMAG is a system for the automated multiple development (AMD) of thin-layer chromatograms. With this system reproducibility gradient evaluation can be employed in TLC, making it possible to separate a substantially larger number of fractions than is feasible with conventional development. It also permits separation of mixtures of substances having markedly different polarities. AMD should find applications in environmental analysis, and in all areas where there are complex samples to be analysed. The system, which is fully automatic, consists of a developing unit and developing chamber, gradient mixer, solvent storage bottles, all valves, switching and setting components, and a separate micro-processor control unit.

Stand D10-11, Hall 6-I

Reader Service No. 116.

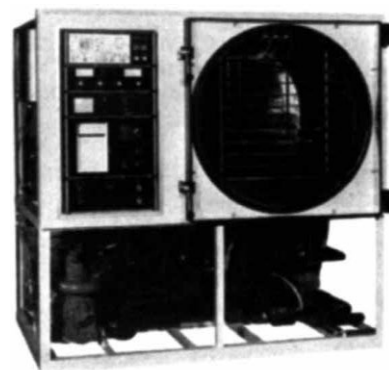


CAMAG's system for TLC analysis.

● The speciality chemicals manufacturer Johnson Matthey Chemicals is exhibiting an extensive range of chemicals at Achema. Five special display modules will be devoted to the company's Honeycat air pollution control units which control the emission of noxious fumes and odours resulting from many manufacturing processes. The four electronics materials modules will be devoted to ultra-high purity raw materials, and metallizing preparations for the electronics industry. The latest addition to the microbonding adhesives range is a conductive silver-glass paste called AuSub (gold sub), the first inorganic adhesive for high reliability die attachment. AuSub has been designed specifically to replace silicon gold eutectics in hermetic packages. Puratronic elements and compounds are a comprehensive range of high purity materials developed to meet the exacting specifications required by the electronics industry. Typical examples are arsenic, gallium and indium for use in III-V semiconductor crystal growth, bismuth germanate and the tungstate systems for compound oxide crystal growth, and boron trioxide and indium for epitaxy applications. The REaction products, comprising all the rare earth elements, are available in a variety of forms and serve as a useful adjunct to the high purity Puratronic series.

Stands D15-20 & E15-20, Hall 5-I

Reader Service No. 117.



The latest Martin Christ freeze-dryer.

● The latest freeze-dryer from Martin Christ GmbH is a unit in which the ice condenser is integral with the drying chamber. During the drying process the water vapour is frozen, taking the shortest path over the ice condenser surfaces which can be cooled to -75°C , installed on right and left of the product shelves. Using this system the drying time can be reduced by 30% compared with conventional systems in which the drying chamber is separate from the ice condenser compartment. The product shelves, up to 4 m^2 in area, are between -50°C and $+70^{\circ}\text{C}$ in a stepless gradient maintained by circulation of a tempering liquid. So thermolabile substances can be both frozen and freeze-dried successfully. The new freeze-dryer is available as either a stand-alone cabinet or for wall installation.

Stand D11, Hall 6-II

Reader Service No. 118.

● Specac's stand at Achema features a wide range of optical components for laser and optical applications. The high performance Fizeau optical test interferometer (FOTI-100) is designed for measuring optical and mechanical flatness and sphericity. It is being demonstrated with its automatic fringe analysis program. The Fizeau interferometer is a versatile topographic tool, consisting of a well-collimated laser beam which is reflected back from both the surface under test and a reference surface. The two reflected wavefronts are superimposed upon a video detector which then displays the resultant fringes on a high-resolution high-linearity video monitor with push-button magnification. The FOTI-100 has a NeNe laser source (wavelength 632.4 μm) and measures flatness to $\lambda/20$ and sphericity to $\lambda/10$. It has an aperture of 100 mm but can be fitted with a beam expander to 150 mm or a beam reducer to 33 mm. A simple two-spot alignment system is used and a hard copy of the displayed interferogram on the screen can be obtained using a Polaroid camera. A wide range of accessories for the FOTI-100 are available including 3 and 5 axis interferometric mounts fitted with a self-centering sample holder. These mounts enable firm but gentle holding of sample and precise positioning which are essential in interferometric testing. A linear or non-linear sliding table together with a horizontal test accessory make for simple measurement of fragile and stress free samples.

Reader Service No. 119.



The TRIO system from Trivector Systems International for chromatography data handling.

● Achema sees the European launch of the Trivector Systems computing integrator, TRIO. Hardware and software are designed with the chromatographer in mind. TRIO offers one or two channel, synchronous or asynchronous data collection from any chromatograph through a high accuracy V/F converter. Incoming chromatograms can be shown in real time on the graphics display. TRIO shows great flexibility in scrolling and magnifying data in real time or post-run mode. The post-run display incorporates baseline and peak identification including graphical comparisons. TRIO complements Trivector's TRILAB, TRISTAR and TRINET to give a wide range of equipment for the data processing needs of the modern analytical laboratory.

Reader Service No. 120.

● The French company Biolafitte is showing a selection of fermentation units at Achema. The emphasis will be on two new automatic *in-situ* sterilizable fermenters, one with a capacity of 20 litres, the other 50 litres. Unlike most units of this type, operation is fully automatic. Integral automation in these units prevents handling errors and reduces the need for an operator to a minimum. The technology used guarantees absolute sterility over a very long period and good stability with time. These Biolafitte fermenters are computer-controlled via a network of multiple microprocessors, and the system can also be connected to the customer's computer system. The computer-control systems, like most other components of Biolafitte reactors, are produced in-house by Biolafitte.

Reader Service No. 121.

Advertisement

TRUST



BIOSIS PREVIEWS® for the biological and biomedical research information you need.

BIOSIS PREVIEWS—the world's largest and most comprehensive source of life science information.

Available through more than 20 online and offline vendors worldwide.

For more information, contact BIOSIS User Services, 2100 Arch Street, Philadelphia, PA 19103-1399, or call toll free (800) 523-4806 (USA except AK, HI, PA) or (215) 587-4800 worldwide. Telex 831739.

BIOSIS is a not-for profit organization serving the biological community since 1926. **BIOSIS PREVIEWS** is a registered trademark of BioSciences Information Service.



Graduate path to entrepreneurship

Graduates should be encouraged to join the ranks of the self-employed. Help is available for those who know where to look.

from Richard Pearson

WITH the labour market picking up rapidly from the doldrums of 1981-82, many graduates will be looking again to the large service and revitalized manufacturing employers for jobs. But what about small firms and self-employment?

In the late 1970s the careers services actively sought out opportunities for graduates in small firms. They did this both to widen the range of opportunities open to graduates and to try and help small firms develop by encouraging them to value a graduate recruit. With graduate unemployment rising to, and exceeding, 12 per cent in 1982, and scientists experiencing even higher rates of unemployment, there was a growing urgency to find new opportunities for graduates both in terms of the type of work they might undertake and by finding new employers. As a result graduates often displaced less-qualified personnel. The fact that graduate unemployment did not rise as steeply as overall unemployment levels and has subsequently fallen significantly as the economy has expanded is testimony both to the flexibility of graduates and to the fact that it has been the higher-skilled jobs that have benefited most from economic expansion.

Indeed, of the half-million extra jobs created in the 18-month period to the end of 1984, two-thirds were for the self-employed, with many of the rest being for part-time workers.

To what extent have, or might, graduates follow the path to such self-employment?

It has been estimated that only about 500 graduates leave higher education each year to launch their own businesses. On degree courses in business studies and management, one might expect a reasonable proportion of graduates to start up their own business ventures, but this apparently is not the case. In fact, it is far more likely to find would-be entrepreneurs emerging from courses in three-dimensional design and other courses in the art and design field, after which self-employment is a normal and expected outcome.

Nevertheless, many graduates do see themselves as having the necessary skills for entrepreneurship and have career aspirations which include the possibility of running their own businesses. A study by Durham Business School¹ has shown that as many students would like to work in small businesses as in large businesses, and that one in four of those with clear career plans wanted to be self-employed or run

their own businesses. Indeed, one in twelve had already started doing something about it, either making preliminary enquiries, doing basic design work, or developing a prototype or patenting an idea.

As part of the study, 1,000 students were asked to rate themselves against a set of entrepreneurial attributes ranging from the 'desire for power' to 'the ability to work long hours'. The results were largely positive, although they were less sure of their ability to risk-take and, most significantly, to attract finance.

There has been a lengthy debate as to whether entrepreneurs are born or made, and if the latter, then what factors can be brought to bear to encourage and aid more people to follow the route through entrepreneurship into self-employment and starting their own businesses.

A preliminary vacation course on entrepreneurship showed positive results in terms of encouraging students towards self-employment, although a subsequent work-experience programme seemed to offer a dose of reality and reduced the numbers believing they were entrepreneurs. Thus while the report concluded that enterprise attributes need to be encouraged by developing educational programmes of the 'education for enterprise' kind, to increase awareness at the undergraduate level, more specifically, there was a need to develop special training programmes for enterprise or small business management.

To encourage graduates to consider starting their own business and to foster the entrepreneurial spirit, the Scottish Enterprise Foundation, an alliance of business, academic and governmental interests, launched Graduate Enterprise in 1982. Based at the University of Stirling, its aim is to provide a four-month training programme covering topics such as business planning and strategy, marketing and market research, accounts and book-keeping and sources of finance.

Such has been the success of the Stirling scheme that the number of places has steadily increased from an initial 20 to 60 in 1985. Significantly, there are plans to adopt the Stirling idea in a number of English universities — the first, based at Cranfield School of Management and directed jointly with the Manpower Services Commission, will receive its first intake of 40 graduates this summer. This is for six weeks of formal training as part of an 18-month programme of support pro-

vided by Cranfield and the scheme's sponsors — Arthur Anderson, the accountants who will provide financial counselling, and BP, who will undertake product development advice and practical help, with National Westminster Bank providing help and advice on financing problems.

As well as the programme of training, counselling and consultancy, the course participants are also eligible for a £4,000 cash package of allowances during the first 18 months, including a basic allowance of £40 per week during training.

With more institutions starting graduate enterprise initiatives, the number of graduates involved is likely to expand significantly. Not all participants will eventually turn out to be successful. If, however, the business success rate of participants is around the national norm, where the success rate for new business helped by an enterprise trust or similar is believed to be around two-thirds, they will have proved their point — that entrepreneurial skills can be developed and business ideas nourished by an appropriate training programme, and this will be of benefit to both the graduates and society.

There are several incentives for self-employment and starting your own business. First is the contraction of alternatives as the number of opportunities in 'big' business and in the public sector decline. Second, a national policy of favourable publicity and special support schemes. These include low-cost advice, premises and finance, support groups such as enterprise trusts, and training programmes. Finally, the academic institutions themselves are being encouraged to follow a more commercially oriented policy.

Those graduates going into self-employment are still to some extent pioneers and few are likely to hit the jackpot immediately. However, such opportunities do exist and with the continuing constraint on traditional job opportunities, students will be wise to widen their horizons in employment terms; if the number of 'traditional' graduate jobs does shrink in the future, the range of opportunities will surely widen. □

¹ *Entrepreneurship and Higher Education* (Durham University Business School, 1984).

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.